DIE GRUNDLEHREN DER MATHEMATISCHEN
WISSENSCHAFTEN IN EINZELDARSTELLUNGEN

VOLUME 104

KAI LAI CHUNG

MARKOV CHAINS
WITH STATIONARY TRANSITION PROBABILITIES

Die Grundlehren der mathematischen Wissenschaften

in Einzeldarstellungen
mit besonderer Berücksichtigung
der Anwendungsgebiete

Band 104

Herausgegeben von

J. L. Doob · E. B. Dynkin · E. Heinz · F. Hirzebruch
E. Hopf · H. Hopf · W. Maak · S. Mac Lane · W. Magnus
D. Mumford · M. M. Postnikow · F. K. Schmidt · K. Stein

Geschäftsführende Herausgeber

B. Eckmann und B. L. van der Waerden

510.8
G889
v. 104
1967

Markov Chains

With Stationary Transition Probabilities

Kai Lai Chung

Professor of Mathematics at Stanford University (U. S. A.)

Second Edition

SPRINGER-VERLAG NEW YORK INC. • 1967

Geschäftsführende Herausgeber:

Prof. Dr. B. Eckmann
Eidgenössische Technische Hochschule Zürich

Prof. Dr. B. L. van der Waerden
Mathematisches Institut der Universität Zürich

All rights reserved, especially that of translation into foreign languages
It is also forbidden to reproduce this book, either whole or in part, by photomechanical means
(photostat, microfilm and/or microcard or any other means)
without written permission from the Publishers

© by Springer-Verlag OHG, Berlin · Göttingen · Heidelberg 1960
© by Springer-Verlag, Berlin · Heidelberg 1967

Library of Congress Catalog Card Number 66-25793

Printed in the United States of America

UNIVERSITY OF ARIZONA
LIBRARY

Title No. 5087

To my parents

67/68 - P

Preface to the First Edition

The theory of Markov chains, although a special case of Markov processes, is here developed for its own sake and presented on its own merits. In general, the hypothesis of a denumerable state space, which is the defining hypothesis of what we call a "chain" here, generates more clear-cut questions and demands more precise and definitive answers. For example, the principal limit theorem (§§ I.6, II.10), still the object of research for general Markov processes, is here in its neat final form; and the strong Markov property (§ II.9) is here always applicable. While probability theory has advanced far enough that a degree of sophistication is needed even in the limited context of this book, it is still possible here to keep the proportion of definitions to theorems relatively low.

From the standpoint of the general theory of stochastic processes, a continuous parameter Markov chain appears to be the first essentially discontinuous process that has been studied in some detail. It is common that the sample functions of such a chain have discontinuities worse than jumps, and these baser discontinuities play a central role in the theory, of which the mystery remains to be completely unraveled. In this connection the basic concepts of separability and measurability, which are usually applied only at an early stage of the discussion to establish a certain smoothness of the sample functions, are here applied constantly as indispensable tools. Hence it is hoped that this book may also serve as an illustration of the modern rigorous approach to stochastic processes toward which there is still so much misgiving.

The two parts of the book, dealing respectively with a discrete and a continuous parameter, are almost independent. It was my original intention to write only the second part, preceded by whatever necessary material from the first. As it turned out, I have omitted details of the continuous parameter analogues when they are obvious enough, in order to concentrate in Part II on those topics which have no counterparts in the discrete parameter case, such as the local properties of sample functions and of transition probability functions. It is these topics that make the continuous parameter case a relatively original and still challenging theory.

Markov process is named after A. A. Markov who introduced the concept in 1907 with a discrete parameter and finite number of states. The denumerable case was launched by Kolmogorov in 1936, followed closely by DOEBLIN whose contributions pervade all parts of the Markov theory. Fundamental work on continuous parameter chains was done by DOOB in 1942 and 1945; and in 1951 PAUL LÉVY, with his unique intuition, drew a comprehensive picture of the field. The present work has grown out of efforts to consolidate and continue the pioneering work of these mathematicians. It is natural that I have based the exposition on my own papers, with major revisions and additions; in particular, the first few sections form an expansion of my lecture notes (mimeographed, Columbia University 1951) which have had some circulation. Quite a few new results, by myself and by colleagues subject to my propaganda, have been as it were made to order for this presentation. Historical comments and credit acknowledgements are to be found in the Notes at the end of the sections. But as a rule I do not try to assign priority to fairly obvious results; to do so would be to insult the intelligence of the reader as well as that of the authors involved.

This book presupposes no knowledge of Markov chains but it does assume the elements of general probability theory as given in a modern introductory course. Part I is on about the same mathematical level as FELLER'S *Introduction to probability theory and its applications, vol. 1.* For Part II the reader should know the elementary theory of real functions such as the oft-quoted theorems of DINI, FATOU, FUBINI and LEBESGUE. He should also be ready to consult, if not already familiar with, certain basic measure-theoretic propositions in DOOB'S *Stochastic processes.* An attempt is made to isolate and expose [*sic*] the latter material, rather than to assure the reader that it is useless luxury. The mature reader can read Part II with only occasional references to Part I.

Markov chains have been used a good deal in applied probability and statistics. In these applications one is generally looking for something considerably more specific or rather more general. In the former category belong finite chains, birth-and-death processes, etc.; in the latter belong various models involving a continuous state space subject to some discretization such as queueing problems. It should be clear that such examples cannot be adequately treated here. In general, the practical man in search of ready-made solutions to his own problems will discover in this book, as elsewhere, that mathematicians are more inclined to build fire stations than to put out fires. A more regrettable omission, from my point of view, is that of a discussion of semigroup or resolvent theory which is pertinent to the last few sections of the book. Let us leave it to another treatise by more competent hands.

A book must be ended, but not without a few words about what lies beyond it. First, sporadic remarks on open problems are given in the Notes. Even for a discrete parameter and in the classical vein, a semblance of fullness exists only in the positive-recurrent case. Much less is known in the null-recurrent case, and a serious study of nonrecurrent phenomena has just begun recently. The last is intimately related to an analysis of the discontinuities of continuous parameter sample functions already mentioned. In the terminology of this book, the question can be put as follows: how do the sample curves manage to go to infinity and to come back from there? A satisfactory answer will include a real grasp on the behavior of instantaneous states, but the question is equally exigent even if we confine ourselves to stable states (as in §§ II.17 to 20). This area of investigations has been called the theory of "boundaries" in analogy with classical analysis, but it is perhaps more succinctly described as an intrinsic theory of compactification of the denumerable state space of the Markov chain. There are a number of allusions to this theme scattered throughout the book, but I have refrained from telling an unfinished story. The solicitous voice of a friend has been heard saying that such a new theory would supersede the part of the present treatment touching on the boundary. Presumably and gladly so. Indeed, to use a Chinese expression, why should the azure not be superior to the blue?

Among friends who have read large portions of the manuscript and suggested valuable improvements are J. L. Doob, Henry P. McKean Jr. and G. E. H. Reuter. My own work in the field, much of it appearing here and some of it for the first time, has been supported in part by the Office of Scientific Research of the United States Air Force. To these, and quite a few others who rendered help of one kind or other, I extend my hearty thanks.

January, 1960 K. L. C.

Preface to the Second Edition

In this revised edition I have added some new material as well as making corrections and improvements. In Part I the additions (in §§ 9, 10, 11) are a few results closely related to the original text and illustrative of the method of taboos. In Part II the major additions have to do with the boundary theory; these include "fine topology" in § 11, "Martin boundary" and "entrance law" in § 19. The old Addenda have been expanded into the new § 12, some of the developments there being also germane to the boundary theory. It is hoped that these efforts have now brought the reader right up to the edge of the boundary. A number of selected items have been inserted in the Bibliography as a further guide to the latest literature on the above-mentioned and other topics.

For some of the important revisions I am particularly indebted to STEVEN OREY and DAVID FREEDMAN for their extensive counsel. All other readers who have sent in corrections are also gratefully acknowledged here. Numerous lesser changes are made, including some for the sake of pedagogy and commentary. Personal taste and habit not being stationary in time, I should have liked to make more radical departures such as deleting hundreds of the ω's in the cumbersome notation, but have generally decided to leave well enough alone. Most of the corrections have already been incorporated in the Russian translation which was published in 1964, but many of the additions are new.

The second edition appears at a time when boundary theory (envisaged in this book as a study in depth of the behavior of sample functions in relation to the "infinities") has just begun to take shape. This vital theme, already announced in the preface to the first edition, will no doubt be the most challenging part of the theory to come. I have chosen not to enter into it in detail in the belief that such a development needs more time to mature. In this regard it may be a timely observation that the theory of Markov processes in general state space, which flourished in recent years and has built up a powerful machinery, has had to date little impact on the denumerable [chain] case. This is because the prevailing assumptions allow the sample functions virtually no other discontinuities than jumps — a situation which would make a trite object of a chain. On the other hand, the special theory of Markov chains has yet to adapt its methodology to a broader context suitable for the general state space. Thus there exists at the moment a state of mutual detachment which surely must not be suffered to continue. Future progress in the field looks to a meaningful fusion of these two aspects of the Markovian phenomenon.

October, 1966 K. L. C.

Contents

Part I. Discrete Parameter

§ 1. Fundamental definitions 1
§ 2. Transition probabilities 5
§ 3. Classification of states 12
§ 4. Recurrence 17
§ 5. Criteria and examples 21
§ 6. The main limit theorem 28
§ 7. Various complements 35
§ 8. Repetitive pattern and renewal process 41
§ 9. Taboo probabilities 45
§ 10. The generating function 54
§ 11. The moments of first entrance time distributions 61
§ 12. A random walk example 71
§ 13. System theorems 76
§ 14. Functionals and associated random variables 81
§ 15. Ergodic theorems 90
§ 16. Further limit theorems 99
§ 17. Almost closed and sojourn sets 112

Part II. Continuous Parameter

§ 1. Transition matrix: basic properties 119
§ 2. Standard transition matrix 128
§ 3. Differentiability 134
§ 4. Definitions and measure-theoretic foundations 140
§ 5. The sets of constancy 148
§ 6. Continuity properties of sample functions 157
§ 7. Further specifications of the process 160
§ 8. Optional random variable 165
§ 9. Strong Markov property 172
§ 10. Classification of states 182
§ 11. Taboo probability functions 187
§ 12. Last exit time 197
§ 13. Ratio limit theorems; discrete approximations 212
§ 14. Functionals 224
§ 15. Post-exit process 230
§ 16. Imbedded renewal process 239
§ 17. The two systems of differential equations 245
§ 18. The minimal solution 251
§ 19. The first infinity 257
§ 20. Examples 272

Bibliography 292

Index 298

Part I. Discrete parameter

§ 1. Fundamental definitions

The precise definition of the term "Markov chain" as used in this monograph will be given below. However, the following remarks will help clarify our usage for the benefit of those readers who have had previous contact with the terminology. A Markov process is a special type of stochastic process distinguished by a certain Markov property; a Markov chain is a Markov process with a denumerable (namely, finite or denumerably infinite) number of states. The time parameter may be taken to be the set of nonnegative integers or the set of nonnegative real numbers; accordingly we have the discrete parameter case or the continuous parameter case. The adjective "simple" is sometimes used to qualify our Markov chain, but since we do not discuss "multiple" chains we shall not make the distinction. Furthermore, we shall discuss only Markov chains "with stationary (or temporally homogeneous) transition probabilities" so that the qualifying phrase in quotes will be understood. Finally, our discussion does not differentiate between a finite or a denumerably infinite number of states so that no special treatment is given to the former case.

In Part I we deal with a discrete parameter. Here the requisite foundations can be summarized as follows.

We consider an abstract set Ω, called the *probability space* (or *sample space*), with the generic element ω, called the *elementary event* (or *sample point*); a Borel field $\mathscr{F}$ of subsets of Ω, called *measurable sets* or *events*, including Ω as a member; and a (countably additive) *probability measure* $\boldsymbol{P}$ defined on $\mathscr{F}$. The triple $(\Omega, \mathscr{F}, \boldsymbol{P})$ is called a *probability triple*. A set in $\mathscr{F}$ of probability zero will be called a *null set*; "almost all ω (a.a. ω)" or "almost everywhere (a.e.)" means "all ω except a null set". The pair $(\mathscr{F}, \boldsymbol{P})$ will be assumed to be *complete* in the sense that every subset of a null set belongs to $\mathscr{F}$ (and is a null set). A Borel subfield of $\mathscr{F}$ is said to be *augmented* iff it contains all null sets. Given a Borel subfield there is a unique smallest augmented Borel subfield of $\mathscr{F}$ containing the given one. Unless otherwise specified all Borel fields used below will be assumed to be augmented. A (real) *random variable (in the generalized sense)* is a single-valued function from a set Δ_0 in $\mathscr{F}$ to the closed real line $X=[-\infty, +\infty]$ such that for every real number c the set of ω in Δ_0 for which $x(\omega) \leqq c$ belongs to $\mathscr{F}$. Δ_0 is called the *domain of definition* of x and the set Δ of ω in Δ_0 for which

$|x(\omega)| < \infty$ is called the *domain of finiteness* of x. *If neither domain is specified it will be understood that* $\boldsymbol{P}(\Delta)=1$, *in other words, the random variable is finite-valued (with probability one).* It follows from the definition that if A is any Borel set in X, the set of ω for which $x(\omega) \in A$, to be denoted by $\{\omega\colon x(\omega) \in A\}$, belongs to $\mathscr{F}$. The probability of this set will be denoted by

$$\boldsymbol{P}\{x(\omega) \in A\}.$$

Symbols similar to these which will appear later should be self-explanatory if we note that commas or semicolons are used to denote intersection of sets (conjunction of events); for example

$$\begin{aligned}\boldsymbol{P}\{x_\nu(\omega) \leqq c_\nu,\ 1 \leqq \nu \leqq 2\} &= \boldsymbol{P}\{x_1(\omega) \leqq c_1;\ x_2(\omega) \leqq c_2\}\\ &= \boldsymbol{P}\left\{\bigcap_{\nu=1}^{2} [x_\nu(\omega) \leqq c_\nu]\right\}.\end{aligned}$$

Given a set $\{x_s,\ s \in S\}$ or random variables, the Borel field *generated* by them is the smallest augmented Borel field with respect to which all the random variables in the set are measurable. Particular cases are $\mathscr{F}\{x_s,\ s \gtrless t\}$ where $\gtrless$ is one of the four symbols $<$, $\leqq$, $>$, $\geqq$, and t is a given real number.

The function F defined by

$$F(u) = \boldsymbol{P}\{x(\omega) \leqq u\}$$

for all real u is called the *distribution function (in the generalized sense)* of x. We have thus

$$\lim_{u \to +\infty} F(u) - \lim_{u \to -\infty} F(u) = \boldsymbol{P}(\Delta).$$

Unless the contrary is stated, we shall mean by a *distribution function* or *a (probability) distribution* that of a random variable which is finite with probability one. Then $\lim\limits_{n \to -\infty} F(u) = 0$, $\lim\limits_{u \to +\infty} F(u) = 1$. A random variable x is *discrete* iff there is a denumerable set A such that $\boldsymbol{P}\{x(\omega) \in A\} = 1$; a *possible value* c of x is one such that $\boldsymbol{P}\{x(\omega) = c\} > 0$. All random variables we encounter in Part I are discrete. A function x from Δ_0 in $\mathscr{F}$ to a denumerable set A is a random variable if and only if the set $\{\omega\colon x(\omega) = c\}$ belongs to $\mathscr{F}$ for every c in A. Indeed A could be any abstract denumerable set and we could define an abstract-valued random variable in this way.

If Λ_1 and Λ_2 are two sets in $\mathscr{F}$, the *conditional probability of* Λ_2 *relative to* (or *under the hypothesis of*, or *given*) Λ_1 is defined by

$$\boldsymbol{P}(\Lambda_2 | \Lambda_1) = \frac{\boldsymbol{P}(\Lambda_1 \Lambda_2)}{\boldsymbol{P}(\Lambda_1)}$$

provided that $\boldsymbol{P}(\Lambda_1) > 0$. If $\boldsymbol{P}(\Lambda_1) = 0$, then $\boldsymbol{P}(\Lambda_2 | \Lambda_1)$ is undefined. Undefined conditional probabilities will appear frequently in what follows as a matter of convenience. If one of them is multiplied by a quantity which is equal to 0, the product is taken to be 0. The conditional probability of the set $\{\omega: x_3(\omega)=c_3\}$ relative to the set $\bigcap_{n=1}^{2}\{\omega: x_n(\omega)=c_n\}$ e.g. is denoted by

$$\boldsymbol{P}\{x_3(\omega)=c_3 \mid x_1(\omega)=c_1,\, x_2(\omega)=c_2\}.$$

The random variables $\{x_\nu,\, 1\leqq \nu \leqq n\}$, not necessarily finite-valued, are said to be *independent* in case

$$\boldsymbol{P}\left\{\bigcap_{\nu=1}^{n}[x_\nu(\omega)\leqq c_\nu]\right\}=\prod_{\nu=1}^{n}\boldsymbol{P}\{x_\nu(\omega)\leqq c_\nu\}$$

for arbitrary real (finite) c_ν, $1\leqq \nu\leqq n$. It follows that the same equation holds if the sets $\{\omega: x_\nu(\omega)\leqq c_\nu\}$ are replaced by the more general sets $\{\omega: x_\nu(\omega)\in A_\nu\}$ where the A_ν are Borel sets in X. The sequence $\{x_n,\, n\geqq 1\}$ is *a sequence of independent random variables* in case any finite number of them are independent. The measurable sets Λ_ν, $1\leqq\nu\leqq n$ or $1\leqq \nu<\infty$ are independent in case their indicators are, the *indicator* of a set being the function which equals one on the set and zero elsewhere.

The *(mathematical) expectation* of a random variable x is the abstract Lebesgue-Stieltjes integral

$$\boldsymbol{E}(x)=\int_{\Omega} x(\omega)\,\boldsymbol{P}(d\omega).$$

Frequently we extend this definition to a random variable that assumes one of the two values $+\infty$ or $-\infty$ with positive probability, provided the integral exists, finite or infinite. If $\Lambda\in\mathscr{F}$ and $\boldsymbol{P}(\Lambda)>0$, the *conditional expectation* of x relative to Λ is defined to be

$$\boldsymbol{E}(x\mid\Lambda)=\int_{\Omega} x(\omega)\,\boldsymbol{P}(d\omega\mid\Lambda)=\frac{\int_{\Lambda} x(\omega)\,\boldsymbol{P}(d\omega)}{\boldsymbol{P}(\Lambda)}.$$

In particular if x is discrete with all its possible values in the denumerable set A then

$$\boldsymbol{E}(x\mid\Lambda)=\frac{1}{\boldsymbol{P}(\Lambda)}\sum_{i\in A} i\,\boldsymbol{P}\{\Lambda;\, x(\omega)=i\}$$

provided the series converges absolutely.

Throughout Part I the letters n, m, ν, r, s, t denote nonnegative integers unless otherwise specified.

A *discrete parameter stochastic process* is a sequence of random variables $\{x_n,\, n\geqq 0\}$ defined with respect to a probability triple $(\Omega,\mathscr{F},\boldsymbol{P})$. If all the random variables are discrete, the union $\boldsymbol{I}$ of all the possible values of all x_n is a denumerable set called the *(minimal) state space* of the process and each element of $\boldsymbol{I}$ a *state*. Thus $i \in \boldsymbol{I}$ if and only if there exists an $n \geqq 0$ such that $\boldsymbol{P}\{x_n(\omega)=i\}>0$. We are borrowing from the language of physics where the term "state" refers to that of a material system whose evolution in time is described by the model that is our stochastic process.

A *discrete parameter Markov chain* is a sequence of discrete random variables $\{x_n,\, n \geqq 0\}$ possessing the following property: for any $n \geqq 2$, $0 \leqq t_1 < \cdots < t_n$ and any $i_1, \ldots, i_n$ in the state space $\boldsymbol{I}$ we have

$$
\left\{\begin{aligned}
&\boldsymbol{P}\{x_{t_n}(\omega)=i_n \mid x_{t_1}(\omega)=i_1, \ldots, x_{t_{n-1}}(\omega)=i_{n-1}\}\\
&\qquad =\boldsymbol{P}\{x_{t_n}(\omega)=i_n \mid x_{t_{n-1}}(\omega)=i_{n-1}\}
\end{aligned}\right. \tag{1}
$$

whenever the left member is defined. The condition (1) will be referred to as the *Markov property*. It is equivalent to the following apparently weaker condition: for every $n \geqq 1$,

$$
\begin{aligned}
&\boldsymbol{P}\{x_n(\omega)=i_n \mid x_0(\omega)=i_0, \ldots, x_{n-1}(\omega)=i_{n-1}\}\\
&\qquad =\boldsymbol{P}\{x_n(\omega)=i_n \mid x_{n-1}(\omega)=i_{n-1}\}.
\end{aligned}
$$

The simple proof is omitted. An important consequence of (1) is that for any $n \geqq 2$ and $0 \leqq t_1 < \cdots < t_n < \cdots < t_{n+m}$; and any $i_1, \ldots, i_n, \ldots, i_{n+m}$ in $\boldsymbol{I}$ we have

$$
\left\{\begin{aligned}
&\boldsymbol{P}\{x_{t_\nu}(\omega)=i_\nu,\ n \leqq \nu \leqq n+m \mid x_{t_\nu}(\omega)=i_\nu,\ 1 \leqq \nu \leqq n-1\}\\
&=\boldsymbol{P}\{x_{t_\nu}(\omega)=i_\nu,\ n \leqq \nu \leqq n+m \mid x_{t_{n-1}}(\omega)=i_{n-1}\}.
\end{aligned}\right. \tag{2}
$$

The proof of (2) proceeds by induction on m. For $m=0$ it reduces to (1). Assuming that (2) is true for a certain value of m, we have, using the rules of combining conditional probabilities and (1):

$$
\begin{aligned}
&\boldsymbol{P}\{x_{t_\nu}(\omega)=i_\nu,\ n \leqq \nu \leqq n+m+1 \mid x_{t_\nu}(\omega)=i_\nu,\ 1 \leqq \nu \leqq n-1\}\\
&=\boldsymbol{P}\{x_{t_\nu}(\omega)=i_\nu,\ n \leqq \nu \leqq n+m \mid x_{t_\nu}(\omega)=i_\nu,\ 1 \leqq \nu \leqq n-1\} \times\\
&\times \boldsymbol{P}\{x_{t_{n+m+1}}(\omega)=i_{n+m+1} \mid x_{t_\nu}(\omega)=i_\nu,\ 1 \leqq \nu \leqq n+m\}\\
&=\boldsymbol{P}\{x_{t_\nu}(\omega)=i_\nu,\ n \leqq \nu \leqq n+m \mid x_{t_{n-1}}(\omega)=i_{n-1}\} \times\\
&\times \boldsymbol{P}\{x_{t_{n+m+1}}(\omega)=i_{n+m+1} \mid x_{t_{n+m}}(\omega)=i_{n+m}\}\\
&=\boldsymbol{P}\{x_{t_\nu}(\omega)=i_\nu,\ n \leqq \nu \leqq n+m \mid x_{t_{n-1}}(\omega)=i_{n-1}\} \times\\
&\times \boldsymbol{P}\{x_{t_{n+m+1}}(\omega)=i_{n+m+1} \mid x_{t_\nu}(\omega)=i_\nu,\ n-1 \leqq \nu \leqq n+m\}\\
&=\boldsymbol{P}\{x_{t_\nu}(\omega)=i_\nu,\ n \leqq \nu \leqq n+m+1 \mid x_{t_{n-1}}(\omega)=i_{n-1}\}.
\end{aligned}
$$

Thus (2) is true with m replaced by $m+1$ and the induction is complete.

A well known result in measure theory asserts that the validity of (2) for all m implies that of the more general result: for every $\mathsf{M} \in \mathscr{F}\{x_t,\, t \geqq t_n\}$ we have

$$
\boldsymbol{P}\{\mathsf{M} \mid x_{t_\nu}(\omega)=i_\nu,\ 1 \leqq \nu \leqq n\}=\boldsymbol{P}\{\mathsf{M} \mid x_{t_n}(\omega)=i_n\}. \tag{3}
$$

In this form the Markov property can be verbally enunciated as postulating that "the probability of an event, knowing the states at several previous moments, is the same as that knowing only the last given state". More loosely it can be stated as requiring that "the past should have no influence on the future except through the present". It is not an uncommon mistake to be misled by these descriptive words to believe that (3) remains true if the conditioning events in it are replaced by more general ones, for example to substitute "$\in A_\nu$" for "$=i_\nu$", where A_ν is a subset of $\boldsymbol{I}$. The resulting equation is of course generally false. However, the following extension of (3) is true and trivial: if $\Lambda \in \mathscr{F}\{x_t, t \leqq t_n\}$

$$\boldsymbol{P}\{\mathrm{M} \mid \Lambda;\, x_{t_n}(\omega)=i\}=\boldsymbol{P}\{\mathrm{M} \mid x_{t_n}(\omega)=i\}. \tag{4}$$

§ 2. Transition probabilities

On the probability triple $(\Omega, \mathscr{F}, \boldsymbol{P})$ let an arbitrary sequence of random variables $\{x_n, n \geqq 0\}$ be given. The Borel field generated by this sequence of random variables will be denoted by $\mathscr{F}\{x_n, n \geqq 0\}$ or simply $\mathscr{F}^0$. This Borel field $\mathscr{F}^0$ is in general a subfield of $\mathscr{F}$, but in a discussion which is concerned solely with the sequence $\{x_n, n \geqq 0\}$ only sets in $\mathscr{F}^0$ will occur. In the case where all the x_n are discrete with the state space $\boldsymbol{I}$, the probabilities of all sets in $\mathscr{F}^0$ are completely determined by the *finite-dimensional joint probabilities*

$$\boldsymbol{P}\{x_0(\omega)=i_0,\, x_1(\omega)=i_1,\, \ldots,\, x_n(\omega)=i_n\}$$

for all $n \geqq 0$ and all $i_n \in \boldsymbol{I}$. The above probability can be expressed as a product of conditional probabilities as follows:

$$\boldsymbol{P}\{x_0(\omega)=i_0\} \prod_{t=1}^{n} \boldsymbol{P}\{x_t(\omega)=i_t \mid x_s(\omega)=i_s,\, 0 \leqq s < t\}.$$

This is true for any stochastic process $\{x_n, n \geqq 0\}$ with the state space $\boldsymbol{I}$. Now if the process is a Markov chain then the conditional probabilities above reduce to those of the form

$$\boldsymbol{P}\{x_t(\omega)=i_t \mid x_{t-1}(\omega)=i_{t-1}\}=p(t;\, i_{t-1},\, i_t)$$

which is a function of t, i_{t-1} and i_t. If we put also

$$\boldsymbol{P}\{x_0(\omega)=i\}=p_i,$$

then we have

$$\boldsymbol{P}\{x_t(\omega)=i_t,\, 0 \leqq t \leqq n\}=p_{i_0} \prod_{t=1}^{n} p(t;\, i_{-1},\, i_t).$$

A particularly interesting case arises when the function $p(\cdot;\cdot,\cdot)$ of three arguments reduces to one of the two states only, thus:

$$(1)\qquad \boldsymbol{P}\{x_n(\omega)=j \mid x_{n-1}(\omega)=i\}=p_{ij}$$

for all $n\geqq 1$ and all i, j in $\boldsymbol{I}$ for which the conditional probability is defined. According to the definition of $\boldsymbol{I}$, for each $i\in\boldsymbol{I}$ there is some n for which the probability in (1) is defined for every j. In this case we have then

$$(2)\qquad \boldsymbol{P}\{x_t(\omega)=i_t,\ 0\leqq t\leqq n\}=p_{i_0}\prod_{t=1}^{n} p_{i_{t-1}i_t}.$$

A discrete parameter Markov chain for which (1) holds as stipulated is said to have *stationary transition probabilities* or, more briefly, to be *homogeneous.* The matrix

$$(p_{ij})\qquad i,j\in\boldsymbol{I}$$

(with any fixed enumeration of the elements of $\boldsymbol{I}$) is called its *(one-step) transition matrix* and the distribution $\{p_i,\ i\in\boldsymbol{I}\}$ its *initial distribution.* The probability p_{ij} is the *(one-step) transition probability from i to j.*

Henceforth in Part I a *discrete parameter Markov chain with stationary transition probabilities* will be abbreviated as "M.C.". This is the name of the mathematical object we study in Part I. *It will be denoted by* $\{x_n,\ n\geqq 0\}$ *or more simply by* $\{x_n\}$ *unless otherwise specified, with an understood probability triple* $(\Omega,\mathscr{F},\boldsymbol{P})$ *and a state space* $\boldsymbol{I}$ *containing the states i, j, k, l, h,* etc. *The symbols* p_i, p_{ij} *and other similar ones to be introduced later all refer to this M.C. unless the contrary is stated.*

The initial and transition probabilities of a M.C. obviously satisfy the following relations: for every i and j,

$$(3)\qquad \begin{cases} p_i\geqq 0, & \sum\limits_i p_i=1;\\ p_{ij}\geqq 0, & \sum\limits_j p_{ij}=1;\end{cases}$$

where, as hereafter, *an unspecified summation over states shall range over all states.* We have seen that these probabilities determine all the joint probabilities by virtue of (2), and therefore they determine the probability measure $\boldsymbol{P}$ on the Borel field $\mathscr{F}^0$ generated by the M.C. In this sense the initial and transition probabilities form a complete set of determining quantities for the M.C. The following alternative definition of a M.C. is often convenient: it is a stochastic process $\{x_n,\ n\geqq 0\}$ with discrete random variables for which there exists a denumerable index set $\boldsymbol{I}$, a sequence $\{p_i,\ i\in\boldsymbol{I}\}$ and a matrix (p_{ij}), $i,j\in\boldsymbol{I}$ satisfying (3) such that the finite-dimensional joint probabilities are given by (2).

We shall now prove that if $\{p_i\}$ and (p_{ij}) are arbitrarily given, subject to the conditions in (3), then a M.C. can be constructed to have them in their proper roles. This theorem is of great theoretical importance. It ensures that the M.C. defined above is nonvacuous, indeed that it exists in abundance. Furthermore, it establishes a general correspondence between certain parts of matrix theory and M.C. theory. We shall return to the latter point shortly.

Theorem 1 (Existence Theorem). Let a denumerable set $\boldsymbol{I}$, a sequence $\{p_i\}$ and a matrix (p_{ij}) with $i, j \in \boldsymbol{I}$ be given satisfying the conditions in (3). There exists a probability triple $(\Omega, \mathscr{F}, \boldsymbol{P})$ and a M.C. $\{x_n, n \geqq 0\}$ defined on it with the state space $\boldsymbol{I}$, the initial distribution $\{p_i\}$ and the transition matrix (p_{ij}).

Proof. For the space Ω we take the set of all infinite sequences $\{\xi_n, n \geqq 0\}$ with each ξ_n an element of $\boldsymbol{I}$. Given any $m \geqq 0$ and any $\{i_0, \ldots, i_m\}$ we denote the set of all sequences $\{\xi_n, n \geqq 0\}$ such that $\xi_\nu = i_\nu$ for $0 \leqq \nu \leqq m$ by $Z(i_0, \ldots, i_m)$ and call such a set a *cylinder set*. For the Borel field $\mathscr{F}$ we take the smallest Borel field containing all cylinder sets. We define a set function $\boldsymbol{P}$ on the set of cylinder sets as follows:

$$\boldsymbol{P}\{Z(i_0, \ldots, i_n)\} = p_{i_0} \prod_{t=1}^{n} p_{i_{t-1} i_t}.$$

By the Kolmogorov extension theorem (Kolmogorov [2]), this set function can be extended to be a probability measure on the Borel field $\mathscr{F}$, since the consistency conditions are easily verified by virtue of (3). We take this as the probability measure $\boldsymbol{P}$. If the pair $(\mathscr{F}, \boldsymbol{P})$ is not yet complete we can complete it in the usual way. Now define a sequence of random variables $\{x_n, n \geqq 0\}$ as follows: if ω in Ω is the sequence $\{\xi_n, n \geqq 0\}$ then we set

$$x_n(\omega) = \xi_n$$

for all $n \geqq 0$. In other words $x_n(\cdot)$ is the n-th *coordinate function* of the point ω regarded as a point in an infinite dimensional space. We shall now verify that the process $\{x_n\}$ is a M.C. with the prescribed $\{p_i\}$ as initial distribution and (p_{ij}) as transition matrix. We have by definition

$$\boldsymbol{P}\{x_0(\omega) = i\} = \boldsymbol{P}\{Z(i)\} = p_i;$$

and if $\boldsymbol{P}\{x_n(\omega) = i\} > 0$,

$$\boldsymbol{P}\{x_{n+1}(\omega) = j \mid x_n(\omega) = i\} = \frac{\boldsymbol{P}\{x_n(\omega) = i,\ x_{n+1}(\omega) = j\}}{\boldsymbol{P}\{x_n(\omega) = i\}}$$

$$= \frac{\boldsymbol{P}\Big\{\bigcup\limits_{i_0, \ldots, i_{n-1}} Z(i_0, \ldots, i_{n-1}, i, j)\Big\}}{\boldsymbol{P}\Big\{\bigcup\limits_{i_0, \ldots, i_{n-1}} Z(i_0, \ldots, i_{n-1}, i)\Big\}} = \frac{\sum\limits_{i_0, \ldots, i_{n-1}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i} p_{ij}}{\sum\limits_{i_0, \ldots, i_{n-1}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i}} = p_{ij}.$$

It remains to prove that $\{x_n\}$ is a M.C. We have, if $n \geqq 2$ and $0 \leqq t_1 < \cdots < t_n$, with self-explanatory notation,

$$\begin{aligned}
&\boldsymbol{P}\{x_{t_n}(\omega)=i_{t_n} \mid x_{t_\nu}(\omega)=i_{t_\nu},\ 1 \leqq \nu \leqq n-1\} \\
&\quad = \sum_{j_{t_\nu}=i_{t_\nu}\ (1\leqq\nu\leqq n)} p_{j_0} \prod_{s=1}^{t_n} p_{j_{s-1}j_s} \Big/ \sum_{j_{t_\nu}=i_{t_\nu}\ (1\leqq\nu\leqq n-1)} p_{j_0} \prod_{s=1}^{t_{n-1}} p_{j_{s-1}j_s} \\
&\quad = \sum_{j_{t_\nu}=i_{t_\nu}\ (\nu=n-1,n)} \prod_{s=t_{n-1}+1}^{t_n} p_{j_{s-1}j_s} \\
&\quad = \sum_{j_{t_\nu}=i_{t_\nu}\ (\nu=n-1,n)} p_{j_0} \prod_{s=1}^{t_n} p_{j_{s-1}j_s} \Big/ \sum_{j_{t_{n-1}}=i_{t_{n-1}}} p_{j_0} \prod_{s=1}^{t_{n-1}} p_{j_{s-1}j_s} \\
&\quad = \boldsymbol{P}\{x_{t_n}(\omega)=i_{t_n} \mid x_{t_{n-1}}(\omega)=i_{t_{n-1}}\}.
\end{aligned}$$

Therefore the Markov property (1.1) holds and Theorem 1 is proved. ▯

We remark that a trivial extension of Theorem 1 yields the existence of a discrete parameter Markov chain with pre-assigned initial distribution and *nonstationary* transition probabilities satisfying the obvious conditions.

We now introduce the *n-step transition probabilities* $p_{ij}^{(n)}$. First of all we set

$$p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

The Kronecker symbol δ_{ij} will be used throughout the monograph. Next, we put $p_{ij}^{(1)} = p_{ij}$ and define successively for $n \geqq 1$:

$$p_{ij}^{(n+1)} = \sum_k p_{ik}^{(n)} p_{kj}^{(1)}. \tag{4}$$

This relation is valid for $n=0$. It will now be shown that for every $\nu \geqq 0$,

$$p_{ij}^{(n)} = \boldsymbol{P}\{x_{\nu+n}(\omega)=j \mid x_\nu(\omega)=i\}, \tag{5}$$

thus justifying the name "n-step transition probability". For $n=1$, (5) reduces to (1). Assuming that (5) is true for a certain value of n and all $i, j \in \boldsymbol{I}$, we have by the Markov property,

$$\begin{aligned}
&\boldsymbol{P}\{x_{\nu+n+1}(\omega)=j \mid x_\nu(\omega)=i\} \\
&= \sum_k \boldsymbol{P}\{x_{\nu+n}(\omega)=k,\ x_{\nu+n+1}(\omega)=j \mid x_\nu(\omega)=i\} \\
&= \sum_k \boldsymbol{P}\{x_{\nu+n}(\omega)=k \mid x_\nu(\omega)=i\} \boldsymbol{P}\{x_{\nu+n+1}(\omega)=j \mid x_\nu(\omega)=i,\ x_{\nu+n}(\omega)=k\} \\
&= \sum_k \boldsymbol{P}\{x_{\nu+n}(\omega)=k \mid x_\nu(\omega)=i\} \boldsymbol{P}\{x_{\nu+n+1}(\omega)=j \mid x_{\nu+n}(\omega)=k\} \\
&= \sum_k p_{ik}^{(n)} p_{kj}^{(1)} = p_{ij}^{(n+1)}.
\end{aligned}$$

Hence (5) is proved by induction on n.

The *absolute probability distribution* of x_n is given by

$$p_i^{(n)} = \boldsymbol{P}\{x_n(\omega) = i\} = \sum_j p_j\, p_{ji}^{(n)}, \qquad n \geqq 0,\ i \in \boldsymbol{I}.$$

The equation (4) may be generalized as follows: for $n \geqq 0$, $m \geqq 0$,

$$p_{ij}^{(n+m)} = \sum_k p_{ik}^{(n)}\, p_{kj}^{(m)}. \tag{6}$$

For $m = 0$ this is trivial. Assuming that (6) is true for a certain value of m and all n, i, j, we have

$$\begin{aligned} p_{ij}^{(n+m+1)} &= \sum_k p_{ik}^{(n+m)}\, p_{kj}^{(1)} = \sum_k \Big\{\sum_l p_{il}^{(n)}\, p_{lk}^{(m)}\Big\}\, p_{kj}^{(1)} \\ &= \sum_l p_{il}^{(n)} \Big\{\sum_k p_{lk}^{(m)}\, p_{kj}^{(1)}\Big\} = \sum_l p_{il}^{(n)}\, p_{lj}^{(m+1)}. \end{aligned}$$

Hence (6) is proved by induction on m.

The formula (4) and its generalization (6), the *Chapman-Kolmogorov equations*, suggest the use of matrices. A *stochastic matrix* is a matrix of finite or (denumerably) infinite order with nonnegative elements such that the sum of each row is equal to one. If $A = (a_{ij})$ and $B = (b_{ij})$ are stochastic matrices their product AB is defined to be the matrix (c_{ij}) where

$$c_{ij} = \sum_k a_{ik}\, b_{kj}.$$

This series always converges and it is easy to see that AB is also a stochastic matrix. Now the transition matrix of a M.C. is obviously a stochastic matrix and Theorem 1 implies conversely that any stochastic matrix is the transition matrix of a M.C., indeed with an arbitrarily given initial distribution. Formula (4) is equivalent to the assertion that the *n-step transition matrix* $(p_{ij}^{(n)})$ is the n-th power of the transition matrix:

$$(p_{ij}^{(n)}) = (p_{ij})^n.$$

Formula (6) becomes the law of exponentiation for matrices. We note also that $(p_{ij}^{(0)}) = (\delta_{ij})$ is the ordinary *unit matrix*. It is now plain that matrix theory can be used as a tool in the study of M.C. In fact, the general theory of a class of matrices including stochastic matrices of finite order has been developed by algebraists and its application to the theory of M.C. has been given in several books; see FRÉCHET [1], ROMANOVSKIĬ [1], and SARYMSAKOV [1]. However, the deeper results in the theory of M.C. can hardly be formulated, much less discovered, by matrix theory. In fact, even when some of our subsequent theorems are intelligible in terms of matrices, their proofs often involve highly probabilistic considerations. Matrix theory will not be used in this monograph.

We close this section by discussing an important special class of M.C.'s. Let $\{y_n, n \geqq 1\}$ be a sequence of independent and identically distributed discrete (real-valued) random variables, and y_0 be a discrete random variable independent of the sequence. Denote by $\boldsymbol{J}$ the set of possible values of y_0, by $\boldsymbol{K}$ that of y_1; and put

$$p_k = \boldsymbol{P}\{y_0 = k\}, \qquad k \in \boldsymbol{J};$$

$$q_k = \boldsymbol{P}\{y_1 = k\}, \qquad k \in \boldsymbol{K}.$$

Let $\overline{\boldsymbol{K}}$ be the least additive semigroup containing $\boldsymbol{K}$ and $\boldsymbol{I} = \boldsymbol{J}(+)\overline{\boldsymbol{K}}$ be the "sum set" of $\boldsymbol{J}$ and $\overline{\boldsymbol{K}}$. Now define the stochastic process $\{x_n, n \geqq 0\}$ as follows:

$$x_n = \sum_{\nu=0}^{n} y_\nu. \tag{7}$$

Theorem 2. This process is a M.C. with state space $\boldsymbol{I}$, initial distribution $\{p_k\}$ regarded as on $\boldsymbol{I}$, and transition matrix (p_{ij}) where

$$p_{ij} = \begin{cases} q_{j-i}, & \text{if } j-i \in \boldsymbol{K}, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Proof. Let $i_0 = j_0 \in \boldsymbol{I}$, $j_\nu = i_\nu - i_{\nu-1}$, $\nu \geqq 1$. Then we have by (7), for every m:

$$\boldsymbol{P}\{x_\nu = i_\nu, 0 \leqq \nu \leqq m\} = \boldsymbol{P}\{y_\nu = j_\nu, 0 \leqq \nu \leqq m\} = p_{j_0} \prod_{\nu=1}^{m} \tilde{q}_{j_\nu} = p_{i_0} \prod_{\nu=1}^{m} p_{i_{\nu-1} i_\nu}$$

where $\tilde{q}_k = q_k$ if $k \in \boldsymbol{K}$ and $\tilde{q}_k = 0$ otherwise. This proves the theorem by the alternative definition of M.C. given above after (3). ▯

We shall call such a M.C. a *chain with independent and stationary increments*. More generally, it is clear that the successive partial sums of a sequence of independent discrete random variables form a discrete parameter Markov chain. Thus, Markovian dependence is indeed a generalization of independence.

The transition probabilities $p_{ij}^{(n)}$ of a chain with independent and stationary increments depend as a function of (i, j) only on the difference $j - i$. This property, sometimes referred to as "spatial homogeneity", is characteristic provided that the state space $\boldsymbol{I}$ possesses certain algebraic structures.

Theorem 3. Let $\{x_n, n \geqq 0\}$ be a M.C. with state space $\boldsymbol{I}$ and transition matrix (p_{ij}). Suppose that $\boldsymbol{I}$ is an additive group (of real numbers) and that each p_{ij} is a function of $j - i$ alone. Then the M.C. is a chain with independent and stationary increments.

Proof. For each $k \in \boldsymbol{I}$ we put

$$q_k = p_{i,i+k}$$

for any $i \in \boldsymbol{I}$, the value on the right being independent of i by hypothesis. We have by the group property of $\boldsymbol{I}$:

$$\sum_{k \in \boldsymbol{I}} q_k = \sum_{k \in \boldsymbol{I}} p_{i,\,i+k} = \sum_{j \in \boldsymbol{I}} p_{ij} = 1 .$$

Now let $j_\nu \in \boldsymbol{I}$, and $i_m = \sum_{\nu=0}^{m} j_\nu$; then $i_m \in \boldsymbol{I}$ by the group property. Define the sequence of random variables $\{y_n,\, n \geqq 0\}$ as follows:

$$(9) \qquad y_0 = x_0; \qquad y_n = x_n - x_{n-1}, \qquad n \geqq 1 .$$

We have then for every m:

$$(10) \qquad \boldsymbol{P}\{y_\nu = j_\nu,\, 0 \leqq \nu \leqq m\} = \boldsymbol{P}\{x_\nu = i_\nu,\, 0 \leqq \nu \leqq m\}$$

$$= p_{i_0} \prod_{\nu=1}^{m} p_{i_{\nu-1} i_\nu} = p_{i_0} \prod_{\nu=1}^{m} q_{j_\nu} .$$

Writing j for j_m above and summing over $j_0, \ldots, j_{m-1}$, we obtain for each $m \geqq 1$:

$$\boldsymbol{P}\{y_m = j\} = q_j .$$

Thus the last term in (10) is equal to

$$\prod_{\nu=0}^{m+1} \boldsymbol{P}\{y_\nu = j_\nu\}$$

and the theorem is proved. ▯

Clearly there are chains with independent and stationary increments which do not fall under Theorem 3; for instance if the y_n's in Theorem 2 take only positive values. However, it can be shown that if $y_0 \equiv 0$ and the set of possible values $\boldsymbol{K}$ above is a set of integers containing both positive and negative ones, then $\boldsymbol{I}$ is indeed the group of integers generated by the greatest common divisor of $\boldsymbol{K}$.

It is interesting that Theorem 3 remains valid if $\boldsymbol{I}$ is the set of positive integers. For in this case the transition matrix (p_{ij}) with the indices in the natural ordering must have identical elements in each diagonal $\{(i,j)\colon j - i = k\}$ with $k \geqq 0$. This forces all the elements of the matrix below the main diagonal to be zero, since each row sum is equal to one. It follows easily that the y_n's, $n \geqq 1$, defined in (9) have the common distribution $\{p_{1,\,1+k},\, k \geqq 0\}$.

On the other hand, if $\boldsymbol{I}$ is the set of integers from 1 to r, the same argument shows that (p_{ij}) must be a "cyclical matrix" of order r such that $p_{i+1,\,j} = p_{i,\,j-1}$ provided all indices are identified modulo r. Even in the simplest case where $r = 2$ and the matrix is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, it is easy to verify that the y_n's, $n \geqq 1$, are neither independent nor identically distributed if $0 < \boldsymbol{P}\{y_0 = 1\} < \frac{1}{2}$. In any case the property of spatial homogeneity may be destroyed by a relabeling of the indices, hence it is not a useful notion when an arbitrary set $\boldsymbol{I}$ is concerned.

Finally, Theorem 3 has an obvious extension to any abstract additive countable group, in particular to an integral lattice in Euclidean space. A chain with independent and stationary increments on such a lattice is sometimes called a *random walk*, (but this term has also been used both in a more special and in a more general context.)

Notes on § 1 and § 2. All the general definitions given in § 1 are conventional except the generalization concerning a random variable with a domain of definition. Frequently it is sufficient to set a random variable equal to $\pm\infty$ where it was not properly defined; but this would be inadequate for our purposes (see in particular § 14). It is somewhat surprising, as it tends to show the rudimentary stage the theory of stochastic processes is still in, that such a generalization has not been indispensable even in a book like DOOB's.

Our usage of the term Markov chain as pertaining to a discrete state space, regardless of the parameter set, follows that of DOOB — a usage which is convenient and justifiable but not yet established. For a discussion of general Markov processes see DOOB [4, § II.6]; also LOÈVE [1, § 25.3]. Naturally our definitions and propositions are special cases of the general theory; for example our Theorem 2.1 corresponds to Example 3 in DOOB [4, p. 86]. We could have phrased a number of the statements to include both the discrete and the continuous parameter case, but we prefer clarity to economy.

A full treatment of random walk on a lattice as defined at the end of the section is to be found in SPITZER [1].

§ 3. Classification of states

In this section certain classifications of the states with regard to their basic transition properties will be given. Further properties leading to finer classifications will appear as we proceed.

We say that *i leads to j* and write $i \rightsquigarrow j$ iff there exists a positive m such that $p_{ij}^{(m)} > 0$. Every state leads to some state. We say that i and *j communicate* iff $i \rightsquigarrow j$ and $j \rightsquigarrow i$; we then write $i \leftrightsquigarrow j$. The relation $\rightsquigarrow$ is in general not reflexive nor symmetric, but it is transitive since

$$p_{ik}^{(m+n)} \geqq p_{ij}^{(m)}\, p_{jk}^{(n)}$$

as a consequence of (2.6). The relation $\leftrightsquigarrow$ is thus symmetric and transitive. It is reflexive over the set of states which communicate with some other state, thus dividing them into disjoint subsets called classes such that two states belong to the same class if and only if they communicate. By fiat each state which does not communicate with any other state shall form a class by itself. Thus a *class* is either a set of two or more mutually communicating states or consists of a single state. The class containing the state i is denoted by $C(i)$. We shall use the phrase "*within a class*" to cover a statement which applies to the states belonging to a fixed class, which may be further specified.

A property defined for all states is called a *class property* iff its possession by one state in a class implies its possession by all the states in the same class. Clearly the negation of a class property is also a class property. If π is a class property, and i possesses π, then the class $C(i)$ is said to possess π. Furthermore if the state space consists of only this class, then we shall say that the M.C. possesses the property π. It is a remarkable fact that many interesting properties for a state turn out to be class properties. The class is a closely knit set of states among which there is great solidarity.

A state that communicates with every state it leads to is called *essential*; otherwise *inessential*. Thus i is essential if and only if $i \rightsquigarrow j$ implies $j \rightsquigarrow i$ for every j; i is inessential if and only if there exists a j such that $i \rightsquigarrow j$ but not $j \rightsquigarrow i$.

Theorem 1. An essential state cannot lead to an inessential state.

Proof. Let i be essential and $i \rightsquigarrow j$. Whenever $j \rightsquigarrow k$ we have also $i \rightsquigarrow k$ by transitivity; consequently $k \rightsquigarrow i$ since i is essential; consequently $k \rightsquigarrow j$ by transitivity. Therefore j is essential by definition. ▯

Corollary. The property of being essential (or inessential) is a class property.

If $i \rightsquigarrow i$, the greatest common divisor of the set of positive n such that $p_{ii}^{(n)} > 0$ is called the *period* of i and denoted by d_i. If i does not lead to itself we do not define its period; such a state will be automatically excluded in any discussion involving a period.

Theorem 2. The property of having a period equal to d is a class property.

Proof. Let $i \leftrightsquigarrow j$. Then there exist positive m and n such that $p_{ij}^{(m)} > 0$, $p_{ji}^{(n)} > 0$. We have if $p_{ii}^{(s)} > 0$,

$$p_{jj}^{(n+s+m)} \geqq p_{ji}^{(n)} \, p_{ii}^{(s)} \, p_{ij}^{(m)} > 0.$$

Since $p_{ii}^{(2s)} > 0$ we have also $p_{jj}^{(n+2s+m)} > 0$. It follows that d_j divides $n+2s+m-(n+s+m)=s$. This being true for every s for which $p_{ii}^{(s)} > 0$ we conclude that d_j divides d_i. Since i and j are interchangeable in the above argument we have $d_j = d_i$. Thus all the states in one class have the same period. $\square$

Theorem 3. To every $j \in C(i)$ there corresponds a unique residue class r_j modulo d_i such that $p_{ij}^{(n)} > 0$ implies that $n \equiv r_j \,(\mathrm{mod}\, d_i)$. Furthermore there exists an $N(j)$ such that $n \geqq N(j)$ implies that $p_{ij}^{(n d_i + r_j)} > 0$.

Proof. Let $p_{ij}^{(m)} > 0$ and $p_{ij}^{(m')} > 0$. There exists an n such that $p_{ji}^{(n)} > 0$. Hence $p_{ii}^{(m+n)} > 0$ and $p_{ii}^{(m'+n)} > 0$. It follows that d_i divides $m - m'$. This proves the first assertion.

By the definition of a period there exist finitely many positive integers n_s, $1 \leqq s \leqq t$, with $p_{ii}^{(n_s)} > 0$ and such that d_i is their greatest common divisor. By an elementary result from number theory there exists an N such that $n \geqq N$ implies the existence of positive integers c_s, $1 \leqq s \leqq t$, satisfying

$$n d_i = \sum_{s=1}^{t} c_s n_s .$$

It follows that if $n \geqq N$, then

$$p_{ii}^{(n d_i)} \geqq \prod_{s=1}^{t} \left(p_{ii}^{(n_s)}\right)^{c_s} > 0 .$$

Now if $j \in C(i)$ and corresponds to the residue class r_j there exists a positive m such that $p_{ij}^{(m d_i + r_j)} > 0$. Let $N(j) = N + m$. If $n \geqq N(j)$ then $n d_i + r_j = n' d_i + m d_i + r_j$ where $n' \geqq N$; hence

$$p_{ij}^{(n d_i + r_j)} \geqq p_{ii}^{(n' d_i)} \, p_{ij}^{(m d_i + r_j)} > 0 .$$

This proves the second assertion. $\square$

The set of states j in $C(i)$ corresponding to the same residue class r modulo d_i as defined in Theorem 3 is called a *subclass* of the class $C(i)$; more particularly it is *the subclass* $C_r(i)$. It is convenient to extend this notation to all r so that

$$C_r(i) = C_s(i) \quad \text{if } r \equiv s \,(\mathrm{mod}\, d_i) .$$

It is clear that the d_i subclasses $C_r(i)$, $1 \leqq r \leqq d_i$, are disjoint and that their union is $C(i)$. Furthermore if $j \in C_r(i)$ then $C_s(j) = C_{r+s}(i)$ for every r and s. Thus the subclasses $C_r(i)$ do not depend on the choice of the state i in the class $C(i)$ except for a cyclic permutation. We shall therefore denote them simply by C_r below when this permutation is immaterial.

Theorem 4. Let C be an essential class of period d. Then

$$C = \bigcup_{r=1}^{d} C_r$$

where the C_r as defined above are disjoint, $C_r=C_s$ if $r\equiv s \pmod d$, and for every $n\geqq 1$:

$$(1) \qquad \sum_{j\in C_{r+n}} p_{ij}^{(n)}=1 \quad \text{if } i\in C_r.$$

Moreover, if $C=\bigcup_{r=1}^{e} D_r$ where the D_r are disjoint, $D_r=D_s$ if $r\equiv s \pmod e$, and for every $n\geqq 1$:

$$\sum_{j\in D_{r+n}} p_{ij}^{(n)}=1 \quad \text{if } i\in D_r;$$

then $d=ef$ where f is a positive integer and each D_r is the union of f of the C_s with the subscript s ranging over a certain residue class modulo e.

Proof. Since C is essential we have by definition and by Theorem 1 $\sum_{j\in C} p_{ij}^{(n)}=1$ for all $n\geqq 1$ if $i\in C$. Hence the first assertion follows immediately from previous remarks. To prove the second assertion, we may suppose, by relabelling the D_r cyclically if necessary, that $C_1D_1\neq 0$. Let $i\in C_1D_1$. Obviously $C_rD_r=C_sD_s$ if $r\equiv s \pmod{de}$. Furthermore we have if $r\geqq 1$

$$(2) \qquad \sum_{j\in C_{r+1}D_{r+1}} p_{ij}^{(r)}=1.$$

Suppose now that for some $r\geqq 1$, $C_rD_r\neq C_r$, then there exists a $j\in C_r-C_rD_r$. From (2) it follows that

$$p_{ij}^{(nde+r-1)}=0$$

for all n. But since $j\in C_r$ this contradicts the second assertion in Theorem 3. Hence for every r, $D_r\supset C_r$. In particular $D_rD_{r+d}\subset D_rC_{r+d}=D_rC_r\neq 0$. Consequently $D_r=D_{r+d}$ and e divides d. Now by hypothesis $D_r=D_{se+r}$ for every r and s. Hence $D_r\supset \bigcup_{s=1}^{f} C_{se+r}$. But $\bigcup_{r=1}^{e} D_r=C=\bigcup_{r=1}^{e}\bigcup_{s=1}^{f} C_{se+r}$. Consequently $D_r=\bigcup_{s=1}^{f} C_{se+r}$ as was to be proved. ▯

In terms of the M.C. $\{x_n, n\geqq 0\}$, the relation (1) implies that for every $m\geqq 0$ and every $i\in C_0$ for which $\boldsymbol{P}\{x_m(\omega)=i\}>0$, we have

$$\boldsymbol{P}\{x_{m+r}(\omega)\in C_r \text{ for all } r\geqq 1 \mid x_m(\omega)=i\}=1.$$

Suppose now that $\boldsymbol{P}\{x_0(\omega)\in C_0\}=1$, viz. $\sum_{i\in C_0} p_i=1$. Consider the sequence of random variables $\{x_{nd}, n\geqq 0\}$. Clearly this is a M.C. with state space C_0. Its initial distribution is $\{p_i, i\in C_0\}$ and its transition matrix is $(p_{ij}^{(d)})$, $i,j\in C_0$. For this *reduced M.C.* the *reduced state space* C_0 is an essential class of period one. It is thus clear why considerations involving the period can be reduced to the case where it is equal to one.

A nonempty set of states A is called *(stochastically) closed* iff we have

$$\sum_{j\in A} p_{ij}=1$$

for every $i\in A$. It then follows that $\sum_{j\in A} p_{ij}^{(n)}=1$ for every n. A closed set is called *minimal* iff it contains no proper subset that is closed. The state space is closed but not necessarily minimal closed.

Theorem 5. The union of a finite number of inessential classes is not closed. A set is minimal closed if and only if it is an essential class.

Proof. We define a partial ordering among inessential classes as follows: $A_1\rightsquigarrow A_2$ iff there are $i_1\in A_1$ and $i_2\in A_2$ such that $i_1\rightsquigarrow i_2$. This is then true for every $i_1\in A_1$ and $i_2\in A_2$. Clearly this relation is reflexive, anti-symmetric between two distinct inessential classes, and transitive; hence it establishes a partial ordering. Now let $A_1, \ldots, A_m$ be distinct inessential classes. Consider all possible partial orderings among them and let $A_{m_1}\rightsquigarrow A_{m_2}\rightsquigarrow \cdots \rightsquigarrow A_{m_l}$ be one of the longest ordered sequences so that there is no n, $1\leqq n\leqq m$, such that $A_{m_l}\rightsquigarrow A_n$. Let $i\in A_{m_l}$. Since i is inessential there is a j such that $i\rightsquigarrow j$ but not $j\rightsquigarrow i$. By the choice of A_m we have $j\notin \bigcup_{n=1}^{m} A_n$, proving that this union is not closed.

It is clear that an essential class is minimal closed. It follows that if a minimal closed set contains an essential state, then it consists of the corresponding class. Now consider a closed set C containing only inessential states. Let i be one of them and let A be the set of all j in C such that $i\rightsquigarrow j$ but not $j\rightsquigarrow i$. Then $A-\{i\}$ is nonempty, closed and a proper subset of C. Hence C cannot be minimal closed. □

The trivial example where $\boldsymbol{I}$ is the set of nonnegative integers, $p_0=1$, and $p_{i,i+1}=1$ for $i\geqq 0$ shows that (i) the closed set $\boldsymbol{I}$ may be the union of a denumerably infinite number of inessential states (classes); (ii) a closed set may contain no minimal closed set.

Let A be closed, $i\in A$ and $m\geqq 0$. Let $\hat{\Omega}=\{\omega\colon x_m(\omega)=i\}$ and suppose $\boldsymbol{P}(\hat{\Omega})>0$; let

$$\hat{\mathscr{F}}=\{\omega\colon x_m(\omega)=i\}\cap\mathscr{F}\{x_s,\, s\geqq m+1\},$$

namely $\hat{\mathscr{F}}$ is the Borel field of sets each of which is the intersection of $\{\omega\colon x_m(\omega)=i\}$ with a set from the Borel field $\mathscr{F}\{x_s,\, s\geqq m+1\}$ defined in §1; and

$$\hat{\boldsymbol{P}}(\Lambda)=\boldsymbol{P}(\Lambda\,|\,\hat{\Omega}) \quad \text{for } \Lambda\in\hat{\mathscr{F}}.$$

On the probability triple $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{\boldsymbol{P}})$ we define a process $\{\hat{x}_n,\, n\geqq 0\}$ as follows:

$$\hat{x}_n=x_{m+n}, \qquad n\geqq 0.$$

Then $\{\hat{x}_n, n \geqq 0\}$ is a M.C. whose state space is a subset $\hat{A}$ of A with $\hat{A}=A$ if in particular A is minimal; whose initial distribution is given by $p_i=1$ and whose transition matrix is $(\hat{p}_{jk})=(p_{jk})$, $j, k \in \hat{A}$. For *conditional transitions within the closed set* A, *relative to the event* $\{\omega: x_m(\omega)=i\}$, the *reduced M.C.* $\{\hat{x}_n\}$ with the *reduced state space* and *reduced transition matrix* suffices. More generally we may consider transitions within A relative to the event $\{\omega: x_m(\omega) \in A\}$. This explains why in the sequel we shall often assume that $\boldsymbol{I}$ is an essential class.

A M.C. is called *irreducible* iff its state space is minimal closed, thus consisting of one essential class. It is called *indecomposable* iff its state space does not contain two disjoint closed sets.

§ 4. Recurrence

For further classification of the states we must introduce new quantities. To employ a vivid language we shall say that *the M.C.* $\{x_n\}$ *is in the state i at the time n or at the n-th step* to mean the event $\{\omega: x_n(\omega)=i\}$. Similar expressions like *starting, entering, returning, avoiding* etc.; and also temporal phrases like *for the first* (or *last*) *time, at least once* etc. are to be interpreted in the same spirit.

The first new quantity is *the probability that the M.C. will be in j for the first time at the n-th step, given that it starts from i.* This is denoted by

$$(1) \qquad \boldsymbol{P}\{x_{m+\nu}(\omega) \neq j,\ 0<\nu<n;\ x_{m+n}(\omega)=j \mid x_m(\omega)=i\},$$

defined for all $m \geqq 0$ for which $\boldsymbol{P}\{x_m(\omega)=i\}>0$. At least one such m exists since $i \in \boldsymbol{I}$. Owing to the stationarity of the transition probabilities the value of this probability does not depend on m whenever it is defined; in fact it is equal to

$$\sum_{i_\nu \neq j\,(0<\nu<n)} p_{i i_1} p_{i_1 i_2} \cdots p_{i_{n-1} j}.$$

With this understanding we shall henceforth write $m=0$ in (1) in spite of the fact that it may well be undefined for $m=0$. In general, a conditional probability like

$$\boldsymbol{P}\{x_{t_\nu}(\omega) \in A_\nu,\ 1 \leqq \nu \leqq n \mid x_0(\omega)=i\}$$

where $0<t_1<\cdots<t_n$ and each $A_\nu \subset \boldsymbol{I}$, will always be interpreted as the common value of

$$\boldsymbol{P}\{x_{m+t_\nu}(\omega) \in A_\nu,\ 1 \leqq \nu \leqq n \mid x_m(\omega)=i\}$$

whenever defined. Similarly, in the sequel when x_0 figures in a discussion, very frequently we are not thinking of the initial random variable of the M.C., but rather using it as an abbreviated but convenient way of talking about the conditional transitions of M.C. For example the literal interpretation of the statement "if $x_0(\omega)=i$, then $x_n(\omega)=i$ for some $n>0$" is that we have a M.C. $\{x_n, n\geqq 0\}$ such that $p_i=\boldsymbol{P}\{x_0(\omega)=i\}=1$, and the assertion is about such a chain with a completely determined initial distribution concentrated at i. This is a possible interpretation but often too strict for our purposes. Our usual interpretation of the statement in quotes will rather be the following formula, valid whenever the conditional probability is defined:

$$\boldsymbol{P}\{x_{m+n}(\omega)=i \text{ for some } n>0 \mid x_m(\omega)=i\}=1.$$

This is an assertion concerning the transition probabilities and has nothing to do with the initial distribution. Formally speaking, it is a question of the process $\{\hat{x}_n, n\geqq 0\}$ defined at the end of §3, with an arbitrary m and $A=\boldsymbol{I}$.

After these remarks we may now define for $n\geqq 1$:

$$f_{ij}^{(n)}=\boldsymbol{P}\{x_\nu(\omega)\neq j,\ 0<\nu<n;\ x_n(\omega)=j \mid x_0(\omega)=i\}$$

$$f_{ij}^{*}=\sum_{n=1}^{\infty} f_{ij}^{(n)}.$$

The use of an asterisk to denote the sum of an infinite series as shown above will be a general notational device (see §9). Thus f_{ij}^* is the probability that the M.C. will be in the state j at least once, given that it starts from the state i. In symbols:

$$\begin{aligned} f_{ij}^*&=\boldsymbol{P}\{x_n(\omega)=j \text{ for some } n>0 \mid x_0(\omega)=i\}\\ &=\boldsymbol{P}\left\{\bigcup_{n=1}^{\infty}[\omega\colon x_n(\omega)=j]\,\middle|\, x_0(\omega)=i\right\}. \end{aligned}$$

Next, the event

$$\limsup_n \{\omega\colon x_n(\omega)=j\}=\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{\omega\colon x_n(\omega)=j\}$$

is the event that the M.C. will be in j *infinitely often*, namely for infinitely many values of n. It will be denoted by $\{\omega\colon x_n(\omega)=j \text{ i.o.}\}$ where "i.o." stands for "infinitely often" and refers to an understood subscript, here n. Its probability given that $x_0(\omega)=i$ will be denoted by

$$g_{ij}=\boldsymbol{P}\{x_n(\omega)=j \text{ i.o.} \mid x_0(\omega)=i\}.$$

A well known result from measure theory states that

$$g_{ij}=\lim_{m\to\infty}\boldsymbol{P}\{x_n(\omega)=j \text{ for some } n\geqq m \mid x_0=i\}. \tag{2}$$

This can be written also as follows:

$$g_{ij}=\lim_{m\to\infty}\sum_k p_{ik}^{(m)} f_{kj}^*.$$

Theorem 1. $i \rightsquigarrow j$ if and only if $f_{ij}^* > 0$; $i \leftrightsquigarrow j$ if and only if $f_{ij}^* f_{ji}^* > 0$.

Proof. This follows from the obvious inequalities

$$\sup_n p_{ij}^{(n)} \leqq f_{ij}^* \leqq \sum_{n=1}^{\infty} p_{ij}^{(n)}. \quad \square$$

Theorem 2. $g_{ij} = f_{ij}^* g_{jj}$.

Proof. We have

$$\begin{aligned} g_{ij} &= \sum_{n=1}^{\infty} \boldsymbol{P}\{x_\nu(\omega) \neq j,\ 0 < \nu < n;\ x_n(\omega) = j;\ x_s(\omega) = j \\ &\qquad\qquad \text{for infinitely many values of } s > n \mid x_0(\omega) = i\} \\ &= \sum_{n=1}^{\infty} \boldsymbol{P}\{x_\nu(\omega) \neq j,\ 0 < \nu < n;\ x_n(\omega) = j \mid x_0(\omega) = i\} \boldsymbol{P}\{x_s(\omega) = j \\ &\qquad\qquad \text{for infinitely many values of } s > n \mid x_n(\omega) = j\} \\ &= \sum_{n=1}^{\infty} f_{ij}^{(n)} g_{jj} = f_{ij}^* g_{jj}. \quad \square \end{aligned}$$

The argument used above is an extremely useful one in the study of M.C. Its general principle may be enunciated as follows. We seek the probability of an event Λ which implies that $x_n(\omega) = j$ for some n; thus

$$\Lambda = \Lambda \bigcup_{n=1}^{\infty} \{\omega : x_n(\omega) = j\}.$$

Decomposing the union into disjoint sets in the usual way, we may be able to write

$$\Lambda = \bigcup_{n=1}^{\infty} [\Lambda_n \cap \{\omega : x_\nu(\omega) \neq j,\ 0 < \nu < n;\ x_n(\omega) = j\}],$$

where $\Lambda_n \in \mathscr{F}\{x_s,\ s \geqq n\}$. Using the Markov property on each Λ_n, we obtain

$$\begin{aligned} \boldsymbol{P}(\Lambda) &= \sum_{n=1}^{\infty} \boldsymbol{P}\{x_\nu(\omega) \neq j,\ 0 < \nu < n;\ x_n(\omega) = j\} \boldsymbol{P}\{\Lambda_n \mid x_n(\omega) = j\} \\ &= \sum_{n=1}^{\infty} f_{ij}^{(n)} \boldsymbol{P}\{\Lambda_n \mid x_n(\omega) = j\}, \end{aligned}$$

where the last expression may simplify. This method, which amounts to the decomposition of the given event according to the time of the first entrance into the state j, will be referred to as *the method of first entrance* for easy reference.

The state i is called *recurrent* or *nonrecurrent* according as $f_{ii}^* = 1$ or < 1.

Theorem 3. $g_{ij} = f_{ij}^*$ or 0 according as j is recurrent or not.

Proof. Let $m \geqq 1$ and

$$g_{ij}(m) = \boldsymbol{P}\{x_n(\omega) = j \text{ for at least } m \text{ values of } n > 0 \mid x_0(\omega) = i\}.$$

By the method of first entrance, we obtain the recursive relation

$$g_{ij}(m+1)=f_{ij}^{*}\, g_{jj}(m).$$

Since $g_{ij}(1)=f_{ij}^{*}$ we have by induction on m,

$$g_{ij}(m+1)=f_{ij}^{*}(f_{jj}^{*})^{m}.$$

The theorem follows since $\lim_{m\to\infty} g_{ij}(m)=g_{ij}$. □

Corollary. $g_{ii}=1$ or 0 according as i is recurrent or not.

Theorem 4. An inessential state is nonrecurrent.

Proof. For every i, j and $m\geqq 0$ we have

$$g_{ij}=\sum_{k} p_{ik}^{(m)}\, g_{kj}.$$

This follows easily from a decomposition according to the value (state) of x_m. If i is inessential there are l and m such that $p_{il}^{(m)}>0$ and $f_{li}^{*}=0$, *a fortiori* $g_{li}=0$. Consequently we have

$$g_{ii}=\sum_{k} p_{ik}^{(m)}\, g_{ki}\leqq \sum_{k} p_{ik}^{(m)}-p_{il}^{(m)}<1.$$

Hence $g_{ii}=0$ and i is nonrecurrent by the preceding Corollary. □

Theorem 5. If $g_{ii}=1$ and $f_{ij}^{*}>0$, then $g_{ij}=1$.

Proof. We have, for every $N>N'>0$,

$$\begin{aligned}
&\boldsymbol{P}\{x_n(\omega)=i \text{ for some } n\geqq N;\ x_n(\omega)\neq j \text{ for all } n\geqq N' \mid x_0(\omega)=i\}\\
&\leqq \sum_{m=N}^{\infty} \boldsymbol{P}\{x_n(\omega)\neq i \text{ for } N\leqq n<m;\ x_m(\omega)=i;\ x_n(\omega)\neq j \text{ for}\\
&\qquad N'\leqq n\leqq N \mid x_0(\omega)=i\}\,\boldsymbol{P}\{x_n(\omega)\neq j \text{ for } n>m \mid x_m(\omega)=i\}\\
&=\sum_{m=N}^{\infty} \boldsymbol{P}\{x_n(\omega)\neq i \text{ for } N\leqq n<m;\ x_m(\omega)=i;\ x_n(\omega)\neq j \text{ for}\\
&\qquad N'\leqq n\leqq N \mid x_0(\omega)=i\}(1-f_{ij}^{*})\\
&=\boldsymbol{P}\{x_n(\omega)=i \text{ for some } n\geqq N;\ x_n(\omega)\neq j \text{ for}\\
&\qquad N'\leqq n\leqq N \mid x_0(\omega)=i\}(1-f_{ij}^{*}).
\end{aligned}$$

Letting $N\to\infty$ and recalling (2), we obtain

$$\begin{aligned}
&\boldsymbol{P}\{x_n(\omega)=i \text{ i.o.};\ x_n(\omega)\neq j \text{ for all } n\geqq N' \mid x_0(\omega)=i\}\\
&\leqq \boldsymbol{P}\{x_n(\omega)=i \text{ i.o.};\ x_n(\omega)\neq j \text{ for all } n\geqq N' \mid x_0(\omega)=i\}(1-f_{ij}^{*}).
\end{aligned}$$

Since $f_{ij}^{*}>0$ it follows that the left member of this inequality is equal to 0. Letting $N'\to\infty$ in it we obtain

$$\boldsymbol{P}\{x_n(\omega)=i \text{ i.o.};\ x_n(\omega)\neq j \text{ for all sufficiently large } n \mid x_0(\omega)=i\}=0.$$

This is equivalent to

$$\begin{aligned}\boldsymbol{P}\{x_n(\omega)=i \text{ i.o.};\ x_n(\omega)=j \text{ i.o.} \mid x_0(\omega)=i\}\\ =\boldsymbol{P}\{x_n(\omega)=i \text{ i.o.} \mid x_0(\omega)=i\}.\end{aligned}$$

The right member is $g_{ii}=1$ by hypothesis, hence the left member is equal to 1, *a fortiori* $g_{ij}=1$. ▯

Corollary 1. If i is recurrent and $i \rightsquigarrow j$, then j is recurrent. Moreover, $g_{ij}=g_{ji}=1$.

Proof. According to the theorem we have $g_{ij}=1$. Hence by Theorem 3, j is recurrent. By Theorem 4, i is essential and so $j \rightsquigarrow i$. Applying Theorem 5 again with i and j interchanged we conclude that $g_{ji}=1$. ▯

Corollary 2. The property of being recurrent (or nonrecurrent) is a class property.

Notes on § 3 and § 4. Our development follows more or less KOLMOGOROV [3]. FELLER [3] does not distinguish between essential and inessential states and calls a nonrecurrent state *transient*; in his second edition he changes "recurrent" into "persistent". FELLER laments the possible literal interpretation of the word "*unessential*"; nevertheless the distinction between the "essential" and "inessential" is rather essential.

KOLMOGOROV defines classes only for essential states. Our slight extension does no harm and leads to one worthwhile statement, namely Theorem 3.5.

The second part of Theorem 3.4 does not seem to have been stated before. Though simple and dispensable, it belongs there logically.

Theorem 4.5 was given by the author in his lectures in 1950. The proof has the advantage of immediate extension to a general state space. There is an intuitive proof using independent random variables and the Borel-Cantelli lemma, but to convert it into a formal argument would perhaps take longer and it would not be generalizable. Another proof may be given by martingale theory; see CHUNG [3].

§ 5. Criteria and examples

We start with an elementary lemma which will be useful on several occasions. Although it is but one half of the well known theorem on the regularity of Nörlund means we give its proof here.

Lemma A. Let $\{a_n, n \geqq 0\}$ be a sequence of nonnegative numbers not all vanishing. A sufficient condition that

$$\lim_{n\to\infty} \frac{\sum_{\nu=0}^{n} a_\nu b_{n-\nu}}{\sum_{\nu=0}^{n} a_\nu} = \lim_{n\to\infty} b_n \tag{1}$$

whenever the sequence $\{b_n, n \geqq 0\}$ of real numbers has a limit is that

$$\lim_{n\to\infty} \frac{a_n}{\sum_{\nu=0}^{n} a_\nu} = 0. \tag{2}$$

Proof. It follows from (2) that

$$\lim_{n\to\infty} \frac{\sum_{\nu=n-N+1}^{n} a_\nu}{\sum_{\nu=0}^{n} a_\nu} = 0 \tag{3}$$

for every $N > 0$. Suppose first that $\lim_{n\to\infty} b_n = b$ where b is finite. Given any $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that $|b_n - b| \leqq \varepsilon$ if $n \geqq N$. Moreover, there exists B such that $|b_n - b| \leqq B$ for all n. We have then

$$\left|\sum_{\nu=0}^{n} a_\nu (b_{n-\nu} - b)\right| \leqq \left(\sum_{\nu=0}^{n-N} a_\nu\right)\varepsilon + \left(\sum_{\nu=n-N+1}^{n} a_\nu\right) B$$

if $n \geqq N$. Dividing by $\sum_{\nu=0}^{n} a_\nu$, letting $n \to \infty$ and using (3) we obtain

$$\overline{\lim_{n\to\infty}} \left| \frac{\sum_{\nu=0}^{n} a_\nu b_{n-\nu}}{\sum_{\nu=0}^{n} a_\nu} - b \right| \leqq \varepsilon.$$

Since ε is arbitrary, (1) follows.

Now let b be infinite; without loss of generality we may suppose that $b = +\infty$. Given any $M > 0$, there exists an $N = N(M)$ such that $b_n \geqq M$ if $n \geqq N$. We have then

$$\sum_{\nu=0}^{n} a_\nu b_{n-\nu} \geqq \left(\sum_{\nu=0}^{n-N} a_\nu\right) M + \left(\sum_{\nu=n-N+1}^{n} a_\nu\right)\left(\min_{0\leqq n\leqq N} b_n\right).$$

Dividing by $\sum_{\nu=0}^{n} a_\nu$, letting $n \to \infty$ and using (3) we obtain

$$\underline{\lim}_{n\to\infty} \frac{\sum_{\nu=0}^{n} a_\nu b_{n-\nu}}{\sum_{\nu=0}^{n} a_\nu} \geqq M.$$

Since M is arbitrary, (1) follows. □

Remark. The condition (2) is satisfied if either

$$\text{(i)} \quad \Sigma a_n < \infty$$

or

$$\text{(ii)} \quad \Sigma a_n = \infty \quad \text{and} \quad a_n \text{ is bounded.}$$

These are the specific conditions under which Lemma A will be applied. We note further that the denominator in (1) may be replaced e.g. by $\sum_{\nu=0}^{n+1} a_\nu$.

One of the basic formulas concerning recurrence is the following. It will be considerably generalized in §9.

Theorem 2. For every i and j in $\boldsymbol{I}$ and $n \geqq 1$ we have

$$p_{ij}^{(n)} = \sum_{\nu=1}^{n} f_{ij}^{(\nu)} p_{jj}^{(n-\nu)}. \tag{4}$$

This is proved by the method of first entrance (into j).

Theorem 3. For every i and j we have

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} p_{ij}^{(n)}}{\sum_{n=0}^{N} p_{jj}^{(n)}} = f_{ij}^*.$$

Proof. Summing (4) from $n=1$ to $n=N$, we have

$$\sum_{n=1}^{N} p_{ij}^{(n)} = \sum_{n=1}^{N} \sum_{\nu=0}^{n-1} f_{ij}^{(n-\nu)} p_{jj}^{(\nu)} = \sum_{\nu=0}^{N-1} p_{jj}^{(\nu)} \sum_{n=\nu+1}^{N} f_{ij}^{(n-\nu)} = \sum_{\nu=0}^{N-1} p_{jj}^{(\nu)} \sum_{n=1}^{N-\nu} f_{ij}^{(n)}.$$

Applying Lemma A with $n=N$, $a_\nu = p_{jj}^{(\nu)}$, and $b_0 = 0$, $b_\nu = \sum_{n=1}^{\nu} f_{ij}^{(n)}$, we obtain Theorem 3. It is clear that the a_ν satisfy either (i) or (ii) in the Remark to Lemma A. ▯

Theorem 4. The state i is recurrent or nonrecurrent according as the series

$$\sum_{n=0}^{\infty} p_{ii}^{(n)}$$

diverges or converges. In the latter case the sum of the series is equal to $(1-f_{ii}^*)^{-1}$.

Proof. Theorem 3 is valid for $i=j$, hence:

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} p_{ii}^{(n)}}{1 + \sum_{n=1}^{N} p_{ii}^{(n)}} = f_{ii}^*.$$

Both assertions now follow from the definition of recurrence. ▯

Remark. Theorem 4 yields another proof of Corollary 2 to Theorem 4.5. For let i be recurrent and $p_{ij}^{(m)} > 0$, $p_{ji}^{(n)} > 0$. Since

$$p_{jj}^{(n+\nu+m)} \geqq p_{ji}^{(n)} p_{ii}^{(\nu)} p_{ij}^{(m)},$$

the divergence of $\sum_\nu p_{ii}^{(\nu)}$ implies that of $\sum_\nu p_{jj}^{(\nu)}$ and consequently j is recurrent.

Theorem 5. The series

$$\sum_{n=0}^{\infty} p_{ij}^{(n)} \tag{5}$$

converges or diverges according as $g_{ij}=0$ or >0.

Proof. If $g_{ij}=0$ then $f_{ij}^* g_{jj}=0$ by Theorem 4.2. If $f_{ij}^*=0$ then the series (5) vanishes termwise. If $g_{jj}=0$ then j is nonrecurrent and $\sum_n p_{jj}^{(n)}$ converges by Theorem 4; hence (5) converges by Theorem 3.

If $g_{ij}>0$ then $f_{ij}^* \geqq g_{ij}>0$ and $g_{jj}=g_{ij}(f_{ij}^*)^{-1}>0$ by Theorem 4.2. Hence j is recurrent by the Corollary to Theorem 4.3; $\sum_n p_{jj}^{(n)}$ diverges by Theorem 4; and (5) diverges by Theorem 3. ▯

Corollary. If j is nonrecurrent then (5) converges. If j is recurrent then (5) diverges or vanishes termwise according as $i \rightsquigarrow j$ or not.

Remark. The following trivial example shows that Theorem 5 is a *best possible* result. Let $0<\alpha<1$; $p_0=1$; $p_{01}=\alpha$, $p_{02}=1-\alpha$, $p_{11}=1$, $p_{22}=1$. Then $g_{01}=\alpha$ and $\sum_{n=1}^{\infty} p_{01}^{(n)} = \sum_{n=1}^{\infty} \alpha = \infty$.

The following theorem summarizes the main facts regarding recurrence within a class.

Theorem 6. Within a recurrent class $f_{ij}^*=g_{ij}=1$ and the series (5) diverges for every i and j. Within a nonrecurrent class $f_{ii}^*<1$, $g_{ij}=0$ and the series (5) converges for every i and j.

We conclude this section with some examples illustrating the various possibilities. These examples are based on the *Bernoullian scheme* or *classical random walk* in one dimension.

Example 1 (free random walk). The state space $\boldsymbol{I}$ is the set of all integers. The M.C. is a chain with independent and stationary increments (§2) with $p_h=1$ for some $h \in \boldsymbol{I}$ and

$$p_{i,i+1}=p, \quad p_{i,i-1}=1-p$$

for all $i \in \boldsymbol{I}$, where $0 \leqq p \leqq 1$. The cases $p=0$ and $p=1$ are completely trivial and are excluded here and in the following examples. Then all states communicate and form an essential class. The period is obviously equal to 2 with the even and odd integers forming two subclasses. An elementary calculation yields

$$p_{00}^{(2n)} = \binom{2n}{n} p^n (1-p)^n .$$

The following analytical formula will be useful: if $0 < 4x < 1$,

$$\sum_{n=0}^{\infty} \binom{2n}{n} x^n = (1-4x)^{-\frac{1}{2}} . \tag{6}$$

Hence if $0 < 4p(1-p) < 1$, we have

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} = [1-4p+4p^2]^{-\frac{1}{2}} = |1-2p|^{-1} . \tag{7}$$

If $4p(1-p)=1$, that is if $p=\frac{1}{2}$, then by a well known theorem of ABEL (see Theorem 10.1 below), we have

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} = \lim_{x \uparrow \frac{1}{4}} (1-4x)^{-\frac{1}{2}} = \infty .$$

We have thus proved that $\sum_{n=0}^{\infty} p_{00}^{(n)}$ converges or diverges according as $p \neq \frac{1}{2}$ or $p=\frac{1}{2}$. Hence by Theorem 4 the M.C. is recurrent if and only if $p=\frac{1}{2}$. This special case is called the *symmetric* case.

Example 2 (random walk with an absorbing barrier). The state space is the set of nonnegative integers: $p_h=1$ for some $h \geqq 1$; and

$$p_{00}=1, \quad p_{i,i+1}=p, \quad p_{i,i-1}=1-p, \qquad i \geqq 1 .$$

The intuitive picture is to start the free random walk in Example 1 from the state h, to proceed as before until the state 0 *(absorbing barrier)* is reached and then to remain there. Formally, let $\{x_n, n \geqq 0\}$ be the M.C. in Example 1; we define $N(\omega)$ to be the smallest n such that $x_n(\omega)=0$, and $\hat{x}_n(\omega)=x_n(\omega)$ if $n \leqq N(\omega)$, $\hat{x}_n(\omega)=0$ if $n > N(\omega)$. Then $\{\hat{x}_n, n \geqq 0\}$ is the M.C. in question. It is clear that all the states except 0 form an inessential class C of period 2; the state 0 is recurrent and forms a class by itself. It is easy to see that if $i > j > 0$ then the f_{ij}^* in Example 2 is the same as in Example 1, because starting from i the M.C. $\{x_n\}$ must enter j before it can enter 0, and the p_{ij} for $i \geqq 1, j \geqq 1$ are the same in the two examples. Furthermore it is clear that

$$f_{ij}^* = (f_{10}^*)^{i-j} .$$

We now calculate f_{10}^* as follows. We have *for the chain in Example 1:*

$$p_{10}^{(2n+1)} = \binom{2n+1}{n} p^n (1-p)^{n+1}, \qquad p_{10}^{(2n)} = 0 .$$

Hence we obtain using (6), if $p \neq \frac{1}{2}$,

$$(8)\qquad \sum_{n=1}^{\infty} p_{10}^{(n)} = \frac{1}{2p}\left[\sum_{n=0}^{\infty}\binom{2n}{n} p^n (1-p)^n - 1\right] = \frac{1}{2p}\left[\frac{1}{|1-2p|} - 1\right].$$

From (7) and (8) we obtain by Theorem 3,

$$(9)\qquad f_{10}^{*} = \frac{1-|1-2p|}{2p} = 1 \quad \text{or} \quad \frac{1-p}{p}$$

according as $p<\frac{1}{2}$ or $p>\frac{1}{2}$. For $p=\frac{1}{2}$ we know $f_{10}^{*}=1$ by recurrence. Thus if $p \leqq \frac{1}{2}$ we have $f_{ij}^{*}=1$ for $i>j\geqq 0$, although C is inessential and therefore nonrecurrent by Theorem 4.4. This explains why no assertion is made regarding f_{ij}^{*}, $i \neq j$, in a nonrecurrent class. It is clear however that at least one of f_{ij}^{*} and f_{ji}^{*} is less than 1 since $f_{ij}^{*}\, f_{ji}^{*} \leqq f_{ii}^{*} < 1$. If $p>\frac{1}{2}$ we have $f_{i0}^{*} = \left(\frac{1-p}{p}\right)^{i}$. In either case $g_{i0}=f_{i0}^{*}$ and $g_{ij}=0$ for all $j\geqq 1$. These facts imply the following: if $p\leqq\frac{1}{2}$,

$$\boldsymbol{P}\left\{\lim_{n\to\infty} x_n(\omega)=0\right\}=1;$$

if $p>\frac{1}{2}$,

$$\boldsymbol{P}\left\{\lim_{n\to\infty} x_n(\omega)=0\right\}=\left(\frac{1-p}{p}\right)^{h},$$

$$\boldsymbol{P}\left\{\lim_{n\to\infty} x_n(\omega)=\infty\right\}=1-\left(\frac{1-p}{p}\right)^{h}.$$

Example 3 (random walk with a reflecting barrier). The state space is the same as in Example 2; $p_h=1$ for some $h\geqq 0$; and

$$p_{00}=1-p, \quad p_{01}=p, \quad p_{i,i+1}=p, \quad p_{i,i-1}=1-p, \quad i\geqq 1.$$

Thus the free random walk is modified so as to return to the state 0 whenever the state -1 would have been entered. It is customary to imagine a *reflecting barrier* placed at $-\frac{1}{2}$. All the states form an essential class of period 1. The same considerations as in Example 2 lead to the validity of (9) for the new M.C. It follows from Theorem 6 that it is nonrecurrent if $p>\frac{1}{2}$. Furthermore a direct argument yields $f_{01}^{(n)}=(1-p)^{n-1}p$ and consequently $f_{01}^{*}=1$ if $p>0$. It is also clear that $f_{00}^{*}\geqq f_{01}^{*} f_{10}^{*}$. Hence if $p\leqq\frac{1}{2}$ we conclude from these facts and (9) that $f_{00}^{*}=1$; it follows from Theorem 6 that the M.C. is recurrent and all $f_{ij}^{*}=1$.

A slight modification of this example is obtained by replacing the first two equations above by the single one: $p_{01}=1$.

Example 4 (random walk with two absorbing barriers). The state space is the set of integers i: $0\leqq i\leqq l$ where $l\geqq 2$; $p_h=1$ for some h,

$0<h<l$; and

$$p_{i,i+1}=p, \quad p_{i,i-1}=1-p, \quad 0<i<l;$$
$$p_{00}=1, \quad p_{ll}=1.$$

The states except 0 and l form an inessential class; each of the two states 0 and l forms a recurrent class by itself. It is clear that $f_{i0}^*+f_{il}^*=1$ for every i, $0<i<l$ since the states $1, 2, \ldots, l-1$ are nonrecurrent. It will follow from more general results (see §12) later that

$$f_{il}^*=\left[1-\left(\frac{1-p}{p}\right)^i\right]\Big/\left[1-\left(\frac{1-p}{p}\right)^l\right].$$

Example 5 (random walk with two reflecting barriers). The state space is the same as in Example 4; $p_h=1$ for some h, $0\leqq h\leqq l$ where $l\geqq 1$; and

$$p_{i,i+1}=p, \quad p_{i,i-1}=1-p, \quad 0<i<l;$$
$$p_{01}=p, \quad p_{00}=1-p, \quad p_{ll}=p, \quad p_{i,l-1}=1-p.$$

All the states form an essential class. Since the state space is finite, for every ω in Ω the infinite sequence $\{x_n(\omega), n\geqq 0\}$ contains some state an infinite number of times; it follows that $g_{hi}>0$ for a certain i. Indeed we have $\sum_{j\in \boldsymbol{I}} g_{hj}\geqq 1$ whenever $\boldsymbol{I}$ is finite. Then $g_{ii}>0$ by Theorem 4.2 and so $g_{ii}=1$ by the Corollary to Theorem 4.3. Thus i is recurrent and so is the M.C. The reader is asked to see for himself if there is a really simpler proof of this obvious fact.

Example 6 (random walk with an absorbing and a reflecting barrier). The state space is the same as in Example 4; $p_h=1$ for some h, $0<h\leqq l$ where $l\geqq 1$; and

$$p_{i,i+1}=p, \quad p_{i,i-1}=1-p, \quad 0<i<l;$$
$$p_{00}=1, \quad p_{ll}=p, \quad p_{l,l-1}=1-p.$$

The states except 0 form an inessential class; the state 0 is recurrent. Since each $i\neq 0$ is nonrecurrent by Theorem 4.4 we have $g_{hi}=0$ if $i\neq 0$. From this and the fact mentioned in Example 5 that $\sum_{j=0}^{l} g_{hj}\geqq 1$ we conclude that $g_{h0}=1$.

Example 7 (random walk with multiple barriers). The state space is the set of all integers. For every i except a set B of *barriers* we have $p_{i,i+1}=p$, $p_{i,i-1}=1-p$. For $j\in B$ we have either $p_{jj}=1$; or $p_{j,j+1}=p$, $p_{jj}=1-p$; or $p_{j,j-1}=1-p$, $p_{jj}=p$. If the set B and also the nature of each barrier are chosen arbitrarily the resulting random walk scheme furnishes a fairly general kind of M.C. Its study can be effected by juxtaposition of the special cases given in Examples 1–6.

The complete evaluation of f_{ij}^* in the above examples will appear (after simple identification) under a more general discussion in a later example (§12).

Notes. For the theorem on Nörlund means see e.g. HARDY [1, Theorem 16], where the condition (2) is shown also to be necessary; cf. loc. cit., Theorem 10 for the case $b=\pm\infty$. Theorem 3 is due to DOEBLIN [4]. For criteria of recurrence in generalized random walks see CHUNG and FUCHS [1], CHUNG and ORNSTEIN [1].

Some of the random walk problems discussed here are usually treated under the heading of *gambler's ruin;* see e.g. USPENSKY [1, Chapter 8], FELLER [3, Chapter 14].

§ 6. The main limit theorem

We shall give in this section a complete determination of the limit or limits of $p_{ij}^{(n)}$ as $n\to\infty$, for every i and j. If j is nonrecurrent it follows already from the Corollary to Theorem 5.5 that $\lim\limits_{n\to\infty} p_{ij}^{(n)}=0$. Furthermore it is evident from (5.4) that the cruical thing is the limiting behavior of $p_{jj}^{(n)}$. Let therefore $i=j$ in that formula:

$$p_{ii}^{(n)}=\sum_{\nu=1}^{n} f_{ii}^{(\nu)}\, p_{ii}^{(n-\nu)}, \qquad n\geqq 1, \tag{1}$$

and let i be recurrent. In this case we have by definition

$$\sum_{u=1}^{\infty} f_{ii}^{(n)}=1 \tag{2}$$

so that $\{f_{ii}^{(n)}, n\geqq 1\}$ determines the probability distribution of a discrete, finite-valued random variable. This will be called the *recurrence time distribution of the recurrent state i* and its first moment the *mean recurrence time,* to be denoted by

$$m_{ii}=\sum_{\nu=1}^{\infty} \nu\, f_{ii}^{(\nu)}\leqq\infty.$$

The random variables corresponding to this distribution will be introduced in §14, and need not concern us here. For our present purpose it is interesting to note that the following theorem is an analytic consequence of the relation (1) together with $p_{ii}^{(0)}=1$ and the condition (2) only.

Theorem 1. If i is a recurrent state with period d_i, then

$$\lim_{n\to\infty} p_{ii}^{(nd_i)}=\frac{d_i}{m_{ii}}.$$

Here as elsewhere, we define $a/b=0$ if a is finite and $b=\pm\infty$.

Proof. In this proof we shall write p_n, f_n, d, m for $p_{ii}^{(n)}$, $f_{ii}^{(n)}$, d_i, m_{ii} respectively. The basic formula becomes

$$p_0=1; \quad p_n=\sum_{\nu=1}^{n} f_\nu p_{n-\nu}, \quad n\geqq 1.$$

We need the following lemma:

Lemma. The period d is the greatest common divisor of the set of integers n for which $f_n>0$.

Proof of the lemma. If s is the smallest integer for which $f_s>0$ then it is clear that it is also the smallest positive integer for which $p_s>0$. Let d'_N be the greatest common divisor of the positive integers $s\leqq n\leqq N$ for which $f_n>0$; d''_N that of those for which $p_n>0$. Since $f_s=p_s$, we have $d'_s=d''_s$. Suppose that $d'_N=d''_N$; we are going to show that $d'_{N+1}=d''_{N+1}$ so that the lemma will follow by induction on N. We have

$$p_{N+1}=\sum_{\nu=1}^{N} f_\nu p_{N+1-\nu}+f_{N+1}.$$

If $f_{N+1}>0$ then $p_{N+1}>0$ and it follows that $d'_{N+1}=d''_{N+1}$. If $f_{N+1}=p_{N+1}=0$ the same conclusion holds. Now suppose that $f_{N+1}=0$, $p_{N+1}>0$. Then there is some ν, $1\leqq\nu\leqq N$ such that $f_\nu p_{N+1-\nu}>0$. Hence d'_N divides ν, d''_N divides $N+1-\nu$ and consequently either number divides $N+1$ since $d'_N=d''_N$. Thus $d'_{N+1}=d'_N=d''_N=d''_{N+1}$ as was to be proved. This establishes the lemma.

We define for $n\geqq 0$,

$$r_n=\sum_{\nu=n+1}^{\infty} f_\nu.$$

Then $r_0=1$ and

(3)
$$\sum_{n=0}^{\infty} r_n=\sum_{\nu=1}^{\infty} \nu f_\nu=m.$$

Since $r_\nu-r_{\nu-1}=-f_\nu$ we have upon substituting into (1),

$$p_n=-\sum_{\nu=1}^{n}(r_\nu-r_{\nu-1})p_{n-\nu}$$

or

$$\sum_{\nu=1}^{n} r_\nu p_{n-\nu}=\sum_{\nu=0}^{n-1} r_\nu p_{n-1-\nu}.$$

This means that the sum does not depend on the value of n; since $r_0 p_0=1$ we have therefore

(4)
$$\sum_{\nu=0}^{n} r_\nu p_{n-\nu}=1, \quad n\geqq 0.$$

Let

$$\lambda = \overline{\lim_{n\to\infty}}\, p_{nd};$$

there exists a subsequence $\{n_k\}$ such that $\lim_{k\to\infty} p_{n_k d} = \lambda$. Let s be such that $f_s > 0$; then d divides s. We have, using (1) and (2),

$$\begin{aligned}\lambda = \lim_{k\to\infty} p_{n_k d} &= \lim_{k\to\infty} \Big\{ f_s\, p_{n_k d - s} + \sum_{\substack{\nu=1\\ \nu\neq s}}^{n_k d} f_\nu\, p_{n_k d - \nu} \Big\} \\ &\leqq f_s \underline{\lim}_{k\to\infty}\, p_{n_k d - s} + \Big(\sum_{\substack{\nu=1\\ \nu\neq s}}^{\infty} f_\nu\Big) \overline{\lim_{k\to\infty}}\, p_{n_k d - \nu} \\ &\leqq f_s \underline{\lim}_{k\to\infty}\, p_{n_k d - s} + (1 - f_s)\,\lambda.\end{aligned}$$

Thus $\underline{\lim}_{k\to\infty} p_{n_k d - s} \geqq \lambda$ and consequently $\lim_{k\to\infty} p_{n_k d - s} = \lambda$ by the definition of λ. This being true for every s for which $f_s > 0$ and every subsequence $\{n_k\}$ for which $\lim_{k\to\infty} p_{n_k d} = \lambda$ we apply the result a finite number of times to conclude that $\lim_{k\to\infty} p_{n_k d - t} = \lambda$ for any t of the form $\sum_{j=1}^{l} c_j\, s_j$ where the c_j and s_j are positive integers such that $f_{s_j} > 0$, $1 \leqq j \leqq l$. By the lemma there exist s_j with $f_{s_j} > 0$, $1 \leqq j \leqq l$, such that their greatest common divisor is equal to d. An elementary result from number theory which we have already used in the proof of Theorem 3.3 then implies the following: if $s \geqq s_0$ then there exist positive integers c_j such that

$$s\, d = \sum_{j=1}^{l} c_j\, s_j.$$

We have therefore proved that for every $s \geqq s_0$,

$$\lim_{k\to\infty} p_{(n_k - s)\, d} = \lambda.$$

Putting $n = (n_k - s_0)\, d$ in (4) and noting that $p_\nu = 0$ unless d divides ν, we obtain

$$\sum_{\nu=0}^{n_k - s_0} r_{\nu d}\, p_{(n_k - s_0 - \nu)\, d} = 1.$$

Letting $k \to \infty$, we obtain

$$\lambda \sum_{\nu=0}^{\infty} r_{\nu d} = 1$$

provided that the series converges; otherwise λ must vanish so that in either case we have

$$\lambda = \frac{1}{\sum_{\nu=0}^{\infty} r_{\nu d}}.$$

Since $f_\nu = 0$ unless d divides ν we see that $r_{\nu d} = \frac{1}{d} \sum_{j=\nu d}^{\nu d+d-1} r_j$. Hence

$$\sum_{\nu=0}^{\infty} r_{\nu d} = \frac{1}{d} \sum_{\nu=0}^{\infty} r_\nu = \frac{m}{d},$$

and we have proved that $\lambda = d/m$.

The preceding evaluation of $\overline{\lim_{n\to\infty}}\, p_{nd}$ applies, *mutatis mutandis*, to $\underline{\lim}_{n\to\infty}\, p_{nd}$. Theorem 1 is therefore proved. ▯

Corollary. Let i and j belong to the same recurrent class of period d and let $j \in C_r(i)$, $1 \leqq r \leqq d$, according to the notation in §3. Then we have

$$\lim_{n\to\infty} p_{ij}^{(nd+r)} = \frac{d}{m_{jj}}.$$

Proof. Since $p_{jj}^{(\nu)} = 0$ unless d divides ν, we have

$$p_{ij}^{(nd+r)} = \sum_{\nu=0}^{n} f_{ij}^{(\nu d+r)}\, p_{jj}^{(nd-\nu d)}. \tag{5}$$

We have $\sum_{\nu=0}^{\infty} f_{ij}^{(\nu d+r)} = \sum_{\nu=1}^{\infty} f_{ij}^{(\nu)} = f_{ij}^* = 1$. Letting $n \to \infty$ and using Lemma A of §5 we obtain the corollary from the theorem. ▯

A recurrent state i is called *positive* or *null* according as $\lim_{n\to\infty} p_{ii}^{(nd_i)}$ is positive or zero. Thus a recurrent state i is positive if and only if $m_{ii} < \infty$.

Theorem 2. The property of a state being positive (or null) is a class property.

Proof. Let $i \leftrightsquigarrow j$ and $p_{ij}^{(m)} > 0$, $p_{ji}^{(n)} > 0$. If d is the common period we have

$$p_{ii}^{(m+\nu d+n)} \geqq p_{ij}^{(m)}\, p_{jj}^{(\nu d)}\, p_{ji}^{(n)}.$$

Letting $\nu \to \infty$ we obtain Theorem 2 from Theorem 1, since i and j are interchangeable in the argument. ▯

Before we consider the limiting behavior of $p_{ij}^{(n)}$ for arbitrary i and j we shall introduce and study certain new quantities. Let the period of j be d. For every integer r we define

$$f_{ij}^*(r) = \sum_{\substack{n=1 \\ n\equiv r\,(\mathrm{mod}\, d)}}^{\infty} f_{ij}^{(n)}.$$

It is clear that

$$f_{ij}^*(r) = \boldsymbol{P}\{x_n(\omega) = j \text{ for some } n \equiv r\,(\mathrm{mod}\ d),\ n > 0 \mid x_0(\omega) = i\}$$

and that

$$\sum_{r=1}^{d} f_{ij}^*(r) = f_{ij}^*.$$

Furthermore, let $C(\alpha)$ with period d_α, $\alpha \geqq 1$, be an enumeration of all the essential classes. For each α let $C_\beta(\alpha)$, $1 \leqq \beta \leqq d_\alpha$, be the subclasses of $C(\alpha)$. For every i we define

$$f^*(r;\, i,\, C_\beta(\alpha)) = \boldsymbol{P}\{x_n(\omega) \in C_\beta(\alpha) \text{ for some } n \equiv r \,(\mathrm{mod}\ d_\alpha),\ n > 0 \mid x_0(\omega) = i\}$$
$$f^*(i,\, C(\alpha)) = \boldsymbol{P}\{x_n(\omega) \in C(\alpha) \text{ for some } n > 0 \mid x_0(\omega) = i\}.$$

It is easy to see that

$$f^*(i,\, C(\alpha)) = \sum_{\beta=1}^{d_\alpha} f^*(r;\, i,\, C_\beta(\alpha))$$

for every $r \geqq 0$; and that

$$\sum_{\alpha} f^*(i,\, C(\alpha)) \leqq 1.$$

Theorem 3. Let $C(\alpha)$ be recurrent. Then for an arbitrary i,

$$f^*(r+s;\, i,\, C_{\beta+s}(\alpha)) = f^*(r;\, i,\, C_\beta(\alpha))$$

for every r and s;

$$f^*(r;\, i,\, C_\beta(\alpha)) = f_{ij}^*(r)$$

for every r and any $j \in C_\beta(\alpha)$; and

$$f^*(i,\, C(\alpha)) = f_{ij}^*$$

for any $j \in C(\alpha)$.

Proof. Let $j \in C_\beta(\alpha)$. Then clearly we have $f_{ij}^*(r) \leqq f^*(r;\, i,\, C_\beta(\alpha))$. On the other hand if $r \geqq 1$ we have by the method of first entrance (into $C_\beta(\alpha)$),

$$f_{ij}^*(r) = \sum_{n=0}^{\infty} \sum_{k \in C_\beta(\alpha)} \boldsymbol{P}\{x_{nd+r}(\omega) = k;\ x_{\nu d+r}(\omega) \notin C_\beta(\alpha),$$
$$0 \leqq \nu < n \mid x_0(\omega) = i\} f_{kj}^*(d).$$

Since $f_{kj}^*(d) = f_{kj}^* = 1$ for every $k \in C_\beta(\alpha)$, the right member above reduces to $f^*(r;\, i,\, C_\beta(\alpha))$. We have thus proved the second equation in Theorem 3. Next let $k \in C_{\beta+s}(\alpha)$. Then $f_{jk}^*(s) = 1$ and it follows that

$$f_{ik}^*(r+s) \geqq f_{ij}^*(r)\, f_{jk}^*(s) = f_{ij}^*(r).$$

Similarly $f_{ij}^*(r) \geqq f_{ik}^*(r+s)$ and we have proved that $f_{ij}^*(r) = f_{ik}^*(r+s)$. This is equivalent to the first equation in view of the second. The third equation is proved in a similar way. ▯

After this excursion we can now complete our study of the limiting behavior of $p_{ij}^{(n)}$ as $n \to \infty$.

Theorem 4. (I) If j is nonrecurrent then $\sum_n p_{ij}^{(n)} < \infty$ for every i so that $\lim_{n\to\infty} p_{ij}^{(n)} = 0$.

(II) If j is recurrent with period d_j and mean recurrence time m_{jj}, and

(a) if i belongs to a different essential class from j, then $p_{ij}^{(n)} = 0$ for every n;

(b) if i belongs to the same class as j so that $j \in C_r(i)$, then

$$\lim_{n\to\infty} p_{ij}^{(nd_j+r)} = \frac{d_j}{m_{jj}}$$

and

$$p_{ij}^{(n)} = 0 \quad \text{if} \quad n \not\equiv r \pmod{d_j};$$

(c) if i is inessential then for every r,

$$\lim_{n\to\infty} p_{ij}^{(nd_j+r)} = f_{ij}^*(r)\,\frac{d_j}{m_{jj}}$$

where

$$0 \leqq f_{ij}^*(r) \quad \text{and} \quad \sum_{r=1}^{d_j} f_{ij}^*(r) \leqq 1 .$$

In particular if j is null all the limits are 0.

Proof. The only new assertion is part (IIc). This follows at once from (5) by an application of §5, Lemma A and Theorem 1. ▯

Corollary. The Cesaro limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} p_{ij}^{(\nu)} = \pi_{ij}$$

exists for every i and j and

$$\pi_{ij} = \frac{f_{ij}^*}{m_{jj}} \tag{6}$$

provided we define $m_{jj} = \infty$ *in case* j *is nonrecurrent.*

We write in particular $\pi_i = \pi_{ii}$ so that $\pi_i = \frac{1}{m_{ii}}$ or 0 according as i is positive-recurrent or not.

The following sharpening of the Corollary to Theorem 4 will be used later (see the proof of Theorem 16.3).

Theorem 5. If

$$m_{ii}^{(2)} = \sum_{\nu=1}^{\infty} \nu^2 f_{ii}^{(\nu)} < \infty$$

then

$$\lim_{n\to\infty} \left[\sum_{\nu=0}^{n} p_{ii}^{(\nu d_i)} - \frac{(n+1)\,d_i}{m_{ii}} \right] = \frac{1}{2m_{ii}^2} \left[m_{ii}^{(2)} - m_{ii}\,d_i \right].$$

Proof. We use the abbreviations in the proof of Theorem 1 above and set furthermore

$$s_n = \sum_{\nu=n}^{\infty} r_\nu .$$

We have from (4),

$$\sum_{\nu=0}^{n} \sum_{t=0}^{n-\nu} r_{td}\, p_{\nu d} = \sum_{\nu=0}^{n} \sum_{t=0}^{\nu} r_{td}\, p_{(\nu-t)d} = n+1;$$

and consequently by (3),

$$\sum_{\nu=0}^{n} s_{\nu d}\, p_{(n-\nu)d} = \sum_{\nu=0}^{n} p_{\nu d}\, s_{(n-\nu)d} = \sum_{\nu=0}^{n} p_{\nu d} \left[m - \sum_{t=0}^{(n-\nu)d-1} r_t \right]$$

$$= m \sum_{\nu=0}^{n} p_{\nu d} - \left(\sum_{\nu=0}^{n} p_{\nu d} \right) d \sum_{t=0}^{n-\nu-1} r_{td} = m \sum_{\nu=0}^{n} p_{\nu d} - n d .$$

Hence it follows from Lemma A of § 5 and Theorem 1 that

$$\lim_{n\to\infty} \left[\sum_{\nu=0}^{n} p_{\nu d} - \frac{n d}{m} \right] = \frac{1}{m} \lim_{n\to\infty} \sum_{\nu=0}^{n} s_{\nu d}\, p_{(n-\nu)d} = \frac{d}{m^2} \sum_{\nu=0}^{\infty} s_{\nu d} .$$

Now it is easy to easy to verify that

$$d \sum_{\nu=0}^{\infty} s_{\nu d} = \sum_{\nu=0}^{\infty} d^2 (\nu+1)\, r_{\nu d} = \tfrac{1}{2} \sum_{t=1}^{\infty} t(t+1)\, d^2 f_{td} ,$$

whence the theorem. □

Notes. Theorem 1 has often been called the "ergodic theorem" for M.C. In conformity with the usage in ergodic theory we reserve this name for a later theorem (§ 15). The average form given in the Corollary to Theorem 4 is then the integrated version of the genuine ergodic theorem.

Theorem 1 was first proved by KOLMOGOROV [3]. His proof is somewhat longer and ours follows that of ERDÖS, FELLER and POLLARD [1]. Despite the elegance and generality of the present proof, the classical proof due essentially to MARKOV himself for the case of a finite number of states is well worth studying (see e.g. FRÉCHET [1], DOOB [4]). For a new approach to Theorem 1 based on probabilistic rather than analytic principles see OREY [3], which is further transformed into a measure-theoretic one by BLACKWELL and FREEDMAN [1].

Our discussion around Theorem 3 seems to clarify certain points in this "nuisance" part which are frequently slurred over.

Theorem 5 is due to FELLER [2], to which we refer for other forms of improvement of Theorem 1.

§ 7. Various complements

The quantities π_i defined in §6 satisfy a certain system of linear homogeneous equations. This *determining system* is not only of theoretical importance but furnishes a practical way of computing these quantities.

Theorem 1. Let C be an essential class of period d. The only solutions $\{u_i, i\in C\}$ of the system of equations

$$u_i = \sum_{k\in C} u_k p_{ki}, \qquad i\in C \tag{1}$$

such that $\sum_{i\in C} |u_i| < \infty$ are given by

$$u_i = c\,\pi_i$$

where c is a constant. If C is nonrecurrent or recurrent-null then every $\pi_i = 0$. If C is recurrent-positive and if C_r, $1\leqq r\leqq d$, are the subclasses, then we have for every r

$$\sum_{i\in C_r} \pi_i = \frac{1}{d} \tag{2}$$

and

$$\sum_{i\in C} \pi_i = 1\,. \tag{3}$$

Proof. (i) To prove that $\{\pi_i, i\in C\}$ is a solution of (1) we need consider only the case where C is a positive class. We have by Theorem 3.4

$$p_{ii}^{(nd)} = \sum_{k\in C} p_{ik}^{(nd-1)} p_{ki} = \sum_{k\in C_{-1}(i)} p_{ik}^{(nd-1)} p_{ki}\,.$$

Letting $n\to\infty$ we have by the Corollary to Theorem 6.1 and FATOU'S lemma

$$\begin{aligned} d\pi_i = \lim_{n\to\infty} p_{ii}^{(nd)} &\geqq \sum_{k\in C_{-1}(i)} \lim_{n\to\infty} p_{ik}^{(nd-1)} p_{ki} \\ &= \sum_{k\in C_{-1}(i)} d\pi_k p_{ki} = d\sum_{k\in C} \pi_k p_{ki}, \end{aligned}$$

or

$$\pi_i \geqq \sum_{k\in C} \pi_k p_{ki}\,. \tag{4}$$

Furthermore we have for each r,

$$1 = \lim_{n\to\infty} \sum_{j\in C_r(i)} p_{ij}^{(nd+r)} \geqq \sum_{j\in C_r(i)} \lim_{n\to\infty} p_{ij}^{(nd+r)} = \sum_{j\in C_r(i)} d\pi_j\,.$$

Consequently we have

$$\sum_{j\in C_r(i)} \pi_j \leqq \frac{1}{d}$$

and

$$(5) \qquad \sum_{j\in C} \pi_j \leqq 1 .$$

Summing (4) over all $i\in C$ we obtain

$$(6) \qquad \sum_{i\in C} \pi_i \geqq \sum_{i\in C} \sum_{k\in C} \pi_k p_{ki} = \sum_{k\in C} \pi_k \sum_{i\in C} p_{ki} = \sum_{k\in C} \pi_k .$$

The interchange of summations is justified by (5). Comparing (4) and (6) we see that equality must hold in (4):

$$\pi_i = \sum_{k\in C} \pi_k p_{ki} .$$

Thus $\{c\,\pi_i, i\in C\}$ for every constant c is a solution of (1).

(ii) To prove the stated uniqueness of the solution above, let $\{u_i\}$ be a solution satisfying $\sum_{i\in C} |u_i| < \infty$. Iterating (1) and interchanging summations we obtain for every n and r:

$$u_i = \sum_{k\in C_{-r}(i)} u_k p_{ki}^{(nd+r)} .$$

Letting $n\to\infty$ we have by the Corollary to Theorem 6.1,

$$(7) \qquad u_i = \Big(\sum_{k\in C_{-r}(i)} u_k\Big) d\pi_i .$$

If C is null then $\pi_i = 0$ so that $u_i = 0$. If C is positive then each $\pi_i > 0$. Equation (7) being true for every r, we see that the bracketed sum does not depend on r. Hence it does not depend on i either, since $C_{-r}(i)$, $1\leqq r\leqq d$, is simply an enumeration of the subclasses of C. Thus in either case $u_i = c\,\pi_i$ as was to be proved.

(iii) In case C is a positive class it follows from (i) and (ii) that we may replace the u's by the π's in (7). Since $\pi_i > 0$ we may divide through to obtain (2) and consequently (3). ▯

Corollary. If C is a positive class, we have for every i,

$$\sum_{j\in C} \pi_{ij} = f^*(i, C) .$$

In particular the sum is equal to one if $i\in C$.

Proof. This follows from Theorem 6.3, the Corollary to Theorem 6.4 and (3) above. ▯

A M.C. $\{x_n, n\geqq 0\}$ such that

$$p_i^{(n)} = \boldsymbol{P}\{x_n(\omega) = i\} = \boldsymbol{P}\{x_0(\omega) = i\} = p_i$$

for every $n\geqq 0$ and $i\in\boldsymbol{I}$ is called a *stationary M.C.* It follows from the definition and the stationarity of the transition probabilities that the joint distribution of $\{x_{n_1+n}, x_{n_2+n}, \ldots, x_{n_l+n}\}$ for any $\{n_1, n_2, \ldots, n_l\}$ is independent of n so that the process is stationary in the established sense.

Theorem 2. Let $\{x_n, n \geqq 0\}$ be a M.C. with the initial distribution $\{p_i\}$ and transition matrix (p_{ij}); and let the positive classes be $D(\alpha)$, $\alpha \in A$ where A is a denumerable set of indices, $D(\alpha) \neq D(\alpha')$ if $\alpha \neq \alpha'$ and $\bigcup_{\alpha \in A} D(\alpha) = D$. In order that the M.C. be stationary it is necessary and sufficient that there exists a sequence of constants $\{\lambda_\alpha, \alpha \in A\}$ with $\lambda_\alpha \geqq 0$, $\sum_{\alpha \in A} \lambda_\alpha = 1$ such that

$$p_i = \begin{cases} 0 & \text{if } i \notin D; \\ \lambda_\alpha \pi_i & \text{if } i \in D(\alpha), \quad \alpha \in A. \end{cases} \tag{8}$$

Proof. (i) Necessity. If $\{n_n, n \geqq 0\}$ is a stationary M.C., then

$$p_i = p_i^{(n)} = \sum_j p_j \, p_{ji}^{(n)}.$$

If $i \notin D$, we have by Theorem 6.4 $\lim_{n \to \infty} p_{ji}^{(n)} = 0$ for every j. Hence letting $n \to \infty$ in the above equation we obtain $p_i = 0$. If $i \in D(\alpha)$, then since $p_{ji}^{(\nu)} = 0$ if $j \in D - D(\alpha)$ we have by what has just been proved

$$p_i = \sum_{j \in D(\alpha)} p_j \, p_{ji}^{(\nu)}$$

for every ν. It follows that

$$p_i = \lim_{n \to \infty} \sum_{j \in D(\alpha)} p_j \left\{ \frac{1}{n} \sum_{\nu=1}^{n} p_{ji}^{(\nu)} \right\}.$$

We can pass to the limit under the summation sign here because the series on the right side is dominated by $\sum_{j \in D(\alpha)} p_j \leqq 1$. Hence by the Corollary to Theorem 6.4 we obtain

$$p_i = \Big(\sum_{j \in D(\alpha)} p_j \Big) \pi_i.$$

Taking $\lambda_\alpha = \sum_{j \in D(\alpha)} p_j$ we see that (8) is satisfied.

(ii) Sufficiency. Let $\{p_i\}$ be defined by (8). Let $i \notin D$; then if $j \in D$, $p_{ji}^{(n)} = 0$ for every n. On the other hand $p_j = 0$ for every $j \notin D$ by definition. Hence we have for every n,

$$p_i^{(n)} = \sum_j p_j \, p_{ji}^{(n)} = 0 = p_i.$$

Now let $i \in D(\alpha)$ and choose any $j \in D(\alpha)$. We have, since $p_k = 0$ if $k \notin D$ and $p_{ki}^{(n)} = 0$ if $k \in D - D(\alpha)$,

$$p_i^{(n)} = \sum_{k \in D(\alpha)} p_k \, p_{ki}^{(n)} = \lambda_\alpha \sum_{k \in D(\alpha)} \pi_k \, p_{ki}^{(n)}.$$

From Theorem 1 we know that

$$\sum_{k\in D(\alpha)} \pi_k p_{ki} = \pi_i.$$

On account of (3) we can iterate this to obtain

$$\sum_{k\in D(\alpha)} \pi_k p_{ki}^{(n)} = \pi_i.$$

We have therefore

$$p_i^{(n)} = \lambda_\alpha \pi_i = p_i.$$

Thus $\{x_n, n\geqq 0\}$ is a stationary M.C. ☐

If C is a positive class the probability distribution given by $\{\pi_i, i\in C\}$ is called the *stationary absolute distribution* for the class.

Corollary to Theorem 2. Let the state space be a positive class. The M.C. is stationary if and only if its initial distribution is the stationary absolute distribution.

The following theorems throw some light on the situation, both analytically and probabilistically. Let $D(\alpha)$ and D be as in Theorem 2 and let

$$f^*(i, D) = \sum_{\alpha\in A} f^*(i, D(\alpha))$$

where the right side is defined in §6. The quantity $f^*(i, D)$ may be called the *probability of absorption of i into some positive class:*

$$f^*(i, D) = \boldsymbol{P}\{x_n(\omega)\in D \text{ for some } n>0 \mid x_0(\omega)=i\}.$$

Theorem 3. The series

$$\sum_j \left(\frac{1}{n}\sum_{\nu=1}^{n} p_{ij}^{(\nu)}\right)$$

in j converges uniformly with respect to n if and only if $\sum_j \pi_{ij}=1$ or equivalently $f^*(i, D)=1$.

Proof. Let $a_{ij}^{(n)} = \frac{1}{n}\sum_{\nu=1}^{n} p_{ij}^{(\nu)}$; then $\lim_{n\to\infty} a_{ij}^{(n)} = \pi_{ij}$ by the Corollary to Theorem 6.4. Hence if $\sum_j a_{ij}^{(n)}$ converges uniformly with respect to n, we have $\sum_j \pi_{ij}=1$ upon passing to the limit under the summation sign. To prove the converse let J be a finite of $\boldsymbol{I}$. We have

$$\lim_{n\to\infty} \sum_{j\notin J} a_{ij}^{(n)} = 1 - \lim_{n\to\infty} \sum_{j\in J} a_{ij}^{(n)} = 1 - \sum_{j\in J} \pi_{ij}.$$

Since $\sum_j \pi_{ij}=1$, given any $\varepsilon>0$ we can choose $J=J(\varepsilon)$ so that

$$1 - \sum_{j\notin J} \pi_{ij} < \varepsilon.$$

There exists $n_0 = n_0(\varepsilon)$ such that if $n > n_0$,

$$\sum_{j \notin J} a_{ij}^{(n)} < \varepsilon.$$

Moreover the set J may be so chosen that the last inequality is also true for $n \leqq n_0$. This proves the uniform convergence of $\sum_j a_{ij}^{(n)}$ with respect to n.

To prove the alternative form of the condition we recall that $\pi_{ij} = 0$ unless $j \in D(\alpha)$ for some $\alpha \in A$ and then $\pi_{ij} = f_{ij}^*/m_{jj}$. Hence we have by Theorem 6.3 and (3) above,

$$\sum_j \pi_{ij} = \sum_{\alpha \in A} \sum_{j \in D(\alpha)} \frac{f_{ij}^*}{m_{jj}} = \sum_{\alpha \in A} f^*(i, D(\alpha)) \sum_{j \in D(\alpha)} \frac{1}{m_{jj}} = \sum_{\alpha \in A} f^*(i, D(\alpha)).$$

This completes the proof of the theorem. ▯

The M.C. is called *non-dissipative* iff $\sum_j \pi_{ij} = 1$ for every i.

Theorem 4. Within an essential class the series $\sum_j p_{ij}^{(n)}$ converges uniformly with respect to n for every i or for no i according as the class is positive or not.

Proof. If the period of the class C is equal to one the proof is similar to that of Theorem 3, since now $\pi_{ij} = \pi_j$ and $\sum_{j \in C} \pi_j = 1$ if and only if C is positive. Let the period be d; then for every r, $1 \leqq r \leqq d$,

$$\sum_j p_{ij}^{(nd+r)} = \sum_{j \in C_r(i)} p_{ij}^{(nd+r)} = 1$$

where $\lim_{n \to \infty} p_{ij}^{(nd+r)} = d\pi_j$ for every $j \in C_r(i)$ and $\sum_{j \in C_r(i)} d\pi_j = 1$. Hence the same conclusion holds. ▯

Corollary. Within a positive class of period d, we have $\lim_{n \to \infty} p_{ij}^{(nd+r)} = d\pi_j$ uniformly with respect to j in $C_r(i)$.

The probabilistic meaning of Theorem 4 is as follows. If i is in a positive class C, then given any $\varepsilon > 0$ there exists a finite set $J(i, \varepsilon)$ of states in C such that for all $n \geqq 0$,

$$\boldsymbol{P}\{x_n(\omega) \in J(\varepsilon) \mid x_0(\omega) = i\} > 1 - \varepsilon.$$

For a discussion of the following system of equations

$$u_i = \sum_{k \in C} p_{ik} u_k, \quad i \in C \tag{9}$$

which is "conjugate" to the system (1), see Theorem 17.5 and its corollary. The following simple results are related to the question.

Theorem 5. Let C be an arbitrary class, not necessarily essential. Suppose that $u_i \geqq 0$ for each $i \in C$ and that we have for some $j \in C$:

$$u_i \geqq \sum_{k \in C - \{j\}} p_{ik} u_k + p_{ij}, \qquad i \in C. \tag{10}$$

Then

$$u_i \geqq f_{ij}^*, \qquad i \in C. \tag{11}$$

Furthermore, $\{f_{ij}^*, i \in C\}$ is the minimal nonnegative solution to the system of inequalities (10), in fact for it there is equality there.

Proof. For each $\nu \geqq 1$, the following relation holds for any two i and j belonging to the same class:

$$\sum_{k \in C - \{j\}} p_{ik} f_{kj}^{(\nu)} = f_{ij}^{(\nu+1)}. \tag{12}$$

This is clear by definition if the sum is extended over $\boldsymbol{I} - \{j\}$. Now if $k \notin C$ and $p_{ik} > 0$, then $i \rightsquigarrow k$ but $k \not\rightsquigarrow i$; hence $k \not\rightsquigarrow j$ and $f_{kj}^{(\nu)} = 0$ so that such a k contributes nothing to the sum, proving the relation as stated. Since $u_i \geqq 0$ we have from (10)

$$u_i \geqq p_{ij} = f_{ij}.$$

Assuming that for every $i \in C$ and some n we have

$$u_i \geqq \sum_{\nu=1}^{n} f_{ij}^{(\nu)},$$

then upon substituting into (10) and using the relation (12) we obtain

$$u_i \geqq \sum_{k \in C - \{j\}} p_{ik} \sum_{\nu=1}^{n} f_{kj}^{(\nu)} + f_{ij} = \sum_{\nu=1}^{n+1} f_{ij}^{(\nu)}.$$

Hence this inequality is true for every $n \geqq 1$ by induction. Letting $n \to \infty$ we obtain (11). Summing (12) over $\nu \geqq 1$ we see that $u_i = f_{ij}^*$ satisfies (10) with equality. ▯

A function $v(\cdot)$ on C satisfying

$$v_i \geqq \sum_{k \in C} p_{ik} v_k, \qquad i \in C, \tag{13}$$

is called *superregular* with respect to (p_{ij}) or the corresponding $\{x_n\}$ on C. It is called *regular* or *subregular* iff the "$\geqq$" in (13) is replaced by "$=$" or "$\leqq$" respectively.

Theorem 6. Let C be an arbitrary class. If C is recurrent, then every nonnegative superregular function is a constant. If C is nonrecurrent, and contains more than one state, then there exists a nonconstant superregular function which is bounded between 0 and 1.

Proof. Iterating (13), we have for every $n \geqq 1$:

$$v_i \geqq \sum_{k \in C} p_{ik}^{(n)} v_k .$$

It follows that either all v_i are zero or all v_i are positive. In the latter case, choose any j and set $u_i = v_i v_j^{-1}$. Then $\{u_i\}$ is a solution of (10). Hence if C is recurrent, we have by (11),

$$u_i \geqq f_{ij}^* = 1$$

or $v_i \geqq v_j$ for every i. Since j is arbitrary this proves the equality of all v_i. On the other hand if C is nonrecurrent and contains more than one state, fix a j and let $v_i = f_{ij}^*$ if $i \neq j$ and $v_j = 1$. Then (13) is satisfied (with equality), and the f_{ij}^*'s are not all equal to one by a previous remark (see §5, Example 2).

Notes. Theorem 1 (for a positive class of period one) seems to be due to FELLER [3], although KOLMOGOROV [3] proved (2) and (3) without the intervention of (1). It should be noted, however, that it is feasible to state the theorem in the form given here, and not merely for a positive class. Otherwise the determination of a positive convergent solution $\{u_i\}$ would not by itself prove that the class is positive. This would be more than a nuisance in practice, since the only general method of showing that a class is positive is precisely to solve the system (1).

Theorems 2 and 3 were given by the author in his lectures in 1950; see LOÈVE [1]. Theorem 4 was proved by KOLMOGOROV [3] "geometrically" in the probabilistic form given after Theorem 4; he then deduced (2) and (3) from it. This approach has some independent interest. For non-dissipative M.C.'s see MAULDON [1].

Theorem 6 was proved by DOOB [5] by martingale theory; the proof given here is new.

§ 8. Repetitive pattern and renewal process

This section is a digression. Its purpose is to establish the equivalence of several notions in current usage.

Let $\{\Omega, \mathscr{F}, \boldsymbol{P}\}$ be a probability triple and let $\{z_n, n \geqq 1\}$ be a sequence of discrete random variables with state space $\boldsymbol{I}$. Suppose that there is a single-valued function R defined for every finite sequence of states and assuming only the values 0 and 1. We say that *the z_n process possesses the repetitive pattern determined by R*, iff the following conditions are satisfied:

(a) if $R(i_1, \ldots, i_m) = 0$ then $R(i_1, \ldots, i_m, i_{m+1}, \ldots, i_{m+n}) = 0$ if and only if $R(i_{m+1}, \ldots, i_{m+n}) = 0$;

(b) if $R(i_1, \ldots, i_m)=0$ and $R(i_1, \ldots, i_m, i_{m+1}, \ldots, i_{m+n})=0$, then

$$\begin{aligned}&\boldsymbol{P}\{z_\nu(\omega)=i_\nu,\ 1\leqq\nu\leqq n+m\}\\ &=\boldsymbol{P}\{z_\nu(\omega)=i_\nu,\ 1\leqq\nu\leqq m\}\boldsymbol{P}\{z_\nu(\omega)=i_{m+\nu},\ 1\leqq\nu\leqq n\}.\end{aligned}$$

The repetitive pattern is said to occur to the process at time n iff

$$R(z_1(\omega), \ldots, z_n(\omega))=0.$$

It is natural to associate with the $\{z_n\}$ process another process $\{y_n,\ n\geqq 0\}$ defined as follows:

$$y_0(\omega)=0, \quad y_n(\omega)=R(z_1(\omega), \ldots, z_n(\omega)), \quad n\geqq 1.$$

Thus the repetitive pattern occurs to the $\{z_n\}$ process at time n if and only if $y_n(\omega)=0$. It follows from (a) and (b) that we have

$$(1)\quad \begin{cases}\boldsymbol{P}\{y_{m+\nu}(\omega)=1,\ 1\leqq\nu<n;\ y_{m+n}(\omega)=0\,|\,y_\nu(\omega)=\varepsilon_\nu,\\ \qquad\qquad\qquad 0\leqq\nu<m;\ y_m(\omega)=0\}\\ =\boldsymbol{P}\{y_\nu(\omega)=1,\ 1\leqq\nu<n;\ y_n(\omega)=0\}\end{cases}$$

for every $m\geqq 0$, $n\geqq 1$ and every set of $\varepsilon_\nu=0$ or 1 for which the conditional probability is defined. If the $\{y_n\}$ process is a M.C. then the condition (1) is clearly satisfied. It is easy to see that in general (1) is less stringent than the Markov property. We call a sequence of random variables $\{y_n,\ n\geqq 0\}$ with $y_0=0$, assuming only the values 0 and 1 and satisfying the condition (1), a *renewal process*. Thus a repetitive pattern generates a renewal process. On the other hand, the renewal process $\{y_n\}$ possesses a *natural* repetitive pattern determined by $R(\varepsilon_1, \ldots, \varepsilon_m)=\varepsilon_m$. We note that it follows from (1) that

$$(2)\quad \begin{cases}\boldsymbol{P}\{y_{m+\nu}(\omega)=1,\ 1\leqq\nu\leqq n\,|\,y_\nu(\omega)=\varepsilon_\nu,\ 0\leqq\nu<m;\ y_m(\omega)=0\}\\ =\boldsymbol{P}\{y_\nu(\omega)=1,\ 1\leqq\nu\leqq n\}.\end{cases}$$

Let

$$\begin{aligned}p_n&=\boldsymbol{P}\{y_n(\omega)=0\},\\ f_n&=\boldsymbol{P}\{y_\nu(\omega)=1,\ 1\leqq\nu<n;\ y_n(\omega)=0\}, \quad n\geqq 1.\\ f_\infty&=\boldsymbol{P}\{y_\nu(\omega)=1,\ 1\leqq\nu<\infty\}.\end{aligned}$$

It follows from (1) by an induction on n that

$$p_n=\boldsymbol{P}\{y_{m+n}(\omega)=0\,|\,y_\nu(\omega)=\varepsilon_\nu,\ 0\leqq\nu<m;\ y_m(\omega)=0\}.$$

We can now prove, exactly as formula (6.1), the *renewal equation:*

$$(3)\qquad p_0=1, \quad p_n=\sum_{\nu=1}^{n} f_\nu p_{n-\nu}, \quad n\geqq 1.$$

It is thus clear that the analogues of Theorems 5.4 and 6.1 must also hold since their proofs depend only on (6.1). This is a purely algebraic-analytical point of view concerning the numbers $\{p_n\}$ and $\{f_n\}$. However, we can go further and associate a M.C. with the renewal process in such a way that a renewal becomes identifiable with the entrance into a certain fixed state.

We shall first give an equivalent formulation of the renewal process. Define a sequence of random variables (in the generalized sense of §1) $\{\tau_t,\ t \geqq 0\}$ as follows. Let $\tau_0(\omega) \equiv 0$ and $\Delta_0 = \Omega$. Suppose that τ_s has been defined on Δ_s, $0 \leqq s < t$; define Δ_t to be the subset of Δ_{t-1} on which there exists an $n > \tau_{t-1}(\omega)$ such that $y_n(\omega) = 0$, and define $\tau_t(\omega)$ to be the smallest such n on Δ_t and ∞ elsewhere. Furthermore define $\varrho_0 = \tau_0$ on Δ_0, and $\varrho_t = \tau_t - \tau_{t-1}$ on Δ_t, $t \geqq 1$. It follows from (1) that $\boldsymbol{P}(\Delta_t) = (1 - f_\infty)^t$, and if $t_1 < \cdots < t_l$ then

$$\boldsymbol{P}\{\varrho_{t_\nu}(\omega) = n_\nu,\ 1 \leqq \nu \leqq l\} = \prod_{\nu=1}^{l} f_{n_\nu} \tag{4}$$

where the last n_l may be ∞. If $f_\infty = 0$, i.e. $\sum_{n=1}^{\infty} f_n = 1$, then the sequence $\{\varrho_t,\ t \geqq 1\}$ is a sequence of independent and finite-valued random variables with the common distribution $\{f_n,\ n \geqq 1\}$. It is easy to see from the definition of the τ_t that we have

$$p_n = \sum_{t=0}^{\infty} \boldsymbol{P}\{\tau_t(\omega) = n\}. \tag{5}$$

Conversely, given a sequence of random variables $\{\varrho_t,\ t \geqq 0\}$ satisfying (4), and $\tau_0(\omega) \equiv 0$, $\tau_t(\omega) = \sum_{s=0}^{t} \varrho_s(\omega)$ for $t \geqq 1$; we can define a renewal process $\{y_n,\ n \geqq 0\}$ as follows:

$$y_n(\omega) = \begin{cases} 0 & \text{if } n = \tau_t(\omega), \quad t \geqq 0; \\ 1 & \text{otherwise}. \end{cases}$$

For this process the left member of (1) is equal to

$$\frac{\prod_\nu f_{n_\nu} f_n}{\prod_\nu f_{n_\nu}} = f_n$$

where the sequence of positive integers $\{n_\nu\}$ satisfying $\sum_\nu n_\nu = m$ is uniquely determined by the sequence $\{\varepsilon_\nu,\ 0 \leqq \nu \leqq m\}$ of 0 and 1 appearing in (1). Hence the $\{y_n\}$ process just defined has the property (1) and is a renewal process. We have thus established a reciprocal correspondence between the renewal process $\{y_n,\ n \geqq 0\}$ and the sequence $\{\varrho_t,\ t \geqq 0\}$ satisfying (4).

We now define a process $\{x_n,\, n\geqq 0\}$ as follows:

$$x_n(\omega)=n-\tau_t(\omega), \quad \text{if} \quad r_t(\omega)\leqq n<\tau_{t+1}(\omega), \qquad t\geqq 0.$$

In words, $x_n(\omega)$ is equal to the smallest value $i\geqq 0$ such that $y_{n-i}(\omega)=0$; namely $x_n(\omega)$ is the "age" at time n in the renewal jargon. Thus $x_n(\omega)=0$ if and only if $y_n(\omega)=0$, namely when a renewal occurs at time n. Furthermore it is clear that if $x_n(\omega)=i$, then $x_{n+1}(\omega)=0$ or $i+1$ regardless of the values of $x_\nu(\omega)$, $1\leqq\nu<n$. Using (1) and (2), we have

$$\begin{aligned}&\boldsymbol{P}\{x_{n+1}(\omega)=0\,|\,x_\nu(\omega)=i_\nu,\ 1\leqq\nu<n;\ x_n(\omega)=0\}\\ &=\boldsymbol{P}\{y_{n+1}(\omega)=0\,|\,y_\nu(\omega)=\varepsilon_\nu,\ 1\leqq\nu<n;\ y_n(\omega)=0\}=f_1;\end{aligned}$$

and if $i\geqq 1$,

$$\begin{aligned}&\boldsymbol{P}\{x_{n+1}(\omega)=i+1\,|\,x_\nu(\omega)=i_\nu,\ 1\leqq\nu<n;\ x_n(\omega)=i\}\\ &=\boldsymbol{P}\{y_{n+1}(\omega)=1\,|\,y_\nu(\omega)=\varepsilon_\nu,\ 1\leqq\nu<n-i;\ y_{n-i}(\omega)=0;\\ &\qquad\qquad\qquad\qquad y_\nu(\omega)=1,\ n-i<\nu\leqq n\}\\ &=\frac{\boldsymbol{P}\{y_\nu(\omega)=1,\ n-i<\nu\leqq n+1\,|\,y_\nu(\omega)=\varepsilon_\nu,\ 1\leqq\nu<n-i;\ y_{n-i}(\omega)=0\}}{\boldsymbol{P}\{y_\nu(\omega)=1,\ n-i<\nu\leqq n\,|\,y_\nu(\omega)=\varepsilon_\nu,\ 1\leqq\nu<n-i;\ y_{n-i}(\omega)=0\}}\\ &=\frac{\boldsymbol{P}\{y_\nu(\omega)=1,\ 0<\nu\leqq i+1\}}{\boldsymbol{P}\{y_\nu(\omega)=1,\ 0<\nu\leqq i\}}=\frac{1-\sum\limits_{\nu=1}^{i+1}f_\nu}{1-\sum\limits_{\nu=1}^{i}f_\nu}.\end{aligned}$$

Consequently the $\{x_n\}$ process is a M.C. with state space the set of integers $\{0, 1, \ldots, l\}$ where l is the smallest integers or ∞ such that $\sum\limits_{1\leqq\nu\leqq l} f_\nu=1$, and with the following transition probabilities

$$p_{i0}=\frac{f_{i+1}}{1-\sum\limits_{\nu=1}^{i}f_\nu}, \qquad p_{i,i+1}=\frac{1-\sum\limits_{\nu=1}^{i+1}f_\nu}{1-\sum\limits_{\nu=1}^{i}f_\nu}.$$

For this M.C. let us calculate $f_{00}^{(n)}$ in the usual notation. We have $f_{00}^{(1)}=p_{00}^{(1)}=f_1$; and if $n\geqq 2$, we have from the above equations

$$f_{00}^{(n)}=\prod_{i=0}^{n-2}p_{i,i+1}\,p_{n-1,0}=f_n. \tag{6}$$

It follows from the identity of (6.1) and (3) above that $p_{00}^{(n)}=p_n$. Alternatively this equation follows directly from

$$p_{00}^{(n)}=\boldsymbol{P}\{x_n(\omega)=0\,|\,x_0(\omega)=0\}=\boldsymbol{P}\{y_n(\omega)=0\}=p_n$$

and then (6) is a consequence.

We have therefore proved that to any renewal process there corresponds a M.C. defined on the same probability triple such that a renewal occurs at time n if and only if the M.C. is in a certain fixed state 0 at time n. Conversely, let $\{x_n, n \geqq 0\}$ be a M.C. such that $x_0(\omega) \equiv 0$ where 0 is an arbitrary state. For $n \geqq 0$ we define $y_n(\omega) = 0$ or 1 according as $x_n(\omega) = 0$ or not. Then it is clear that $\{y_n, n \geqq 0\}$ is a renewal process. Thus a renewal process, and more specifically a process possessing a repetitive pattern, is equivalent to a M.C. in the sense elaborated above. The renewal or occurrence of the repetitive pattern is identifiable in this equivalence with the entrance of the M.C. into a certain fixed state which may be recurrent or nonrecurrent. The repetitive pattern may be called *recurrent* or *nonrecurrent* accordingly.

Notes. The term "recurrent event" has also been used (FELLER [2], [3]) for "repetitive pattern". It is called "certain" or "uncertain" according as the pattern is recurrent or nonrecurrent in our terminology.

The conditions (a) and (b) for a repetitive pattern are those given by FELLER [3], but as DOOB pointed out (oral communication), it would be more logical to omit the words "and $R(i_1, \ldots, i_m, i_{m+1}, \ldots, i_n) = 0$" in (b). The modified definitions is easily seen to be equivalent to the original one in case $f_\infty = 0$, namely when the repetitive pattern is recurrent. It is, however, more stringent in general.

Our definition of a renewal process may not be orthodox, but the equivalent one in terms of the sequence $\{\varrho_t\}$ has been so named (DOOB [3]). The latter formulation leads to extensions where the ϱ_t need no longer be discrete or positive. In the discrete, but not necessarily positive, case the main limit theorem of §6 may be extended to apply to the p_n defined in (5). This was first considered by CHUNG and WOLFOWITZ [1]. For the general, not necessarily discrete, case see BLACKWELL [2], W. L. SMITH [1] and FELLER [6]. For the literature on renewal theory see the references given in SMITH's paper. From another point of view, the renewal process is the simplest special case of a semi-Markovian process defined by LÉVY and the author; see LÉVY [4].

§ 9. Taboo probabilities

For a deeper study of the M.C. $\{x_n, n \geqq 0\}$ we now introduce *transition probabilities with taboo states*. Let H be an arbitrary set of states. We define

$$ {}_Hp_{ij}^{(n)} = \boldsymbol{P}\{x_n(\omega) = j;\ x_\nu(\omega) \notin H,\ 0 < \nu < n \mid x_0(\omega) = i\}, \qquad n \geqq 1. $$

In words, ${}_Hp_{ij}^{(n)}$ is the probability, starting from the state i, of entering the state j at the n-th step *under the restriction that none of the states in* *the set H is entered in-between* (exclusive of both ends); or again that the transition is to be restricted within the set $\boldsymbol{I}-H$. The set H is called the *taboo set* and the states in it are called *taboo states* for this transition or its corresponding probability. H may be empty; in this case it is omitted from the notation. If H consists of a single state k we shall denote the probability by ${}_kp_{ij}^{(n)}$. If the taboo set consists of the union of a set H and a state k (which may belong to H), the corresponding probability will be denoted by ${}_{k,H}p_{ij}^{(n)}$. We note that in our previous notation, $f_{ij}^{(n)}={}_jp_{ij}^{(n)}$; in the same spirit we shall write ${}_Hf_{ij}^{(n)}={}_{j,H}p_{ij}^{(n)}$. We define ${}_Hp_{ij}^{(0)}=\delta_{ij}$ if $i\notin H$; $=0$ if $i\in H$. Note also that ${}_Hp_{ij}^{(1)}=p_{ij}^{(1)}$. Finally we define

$$ {}_Hp_{ij}^{*}=\sum_{n=1}^{\infty}{}_Hp_{ij}^{(n)}; $$

if the series diverges its value is ∞. This is of course the expected number of times the M.C. is in the state j before it is in the set H, starting from i. The usual conventions regarding the arithmetic of ∞ are adopted; thus

$$ c+\infty=\infty,\qquad |c|\cdot\infty=\infty,\qquad \frac{c}{\infty}=0,\qquad \frac{\infty}{|c|}=\infty,\qquad \infty-c=\infty $$

where $c\neq 0$ is a finite constant;

$$ \infty+\infty=\infty,\qquad \infty\cdot\infty=\infty,\qquad 0\cdot\infty=0; $$

but $\infty-\infty$ or ∞/∞ is a forbidden operation.

We are now going to give two formulas which are the primary sources of all the formulas in this section.

Theorem 1 (Basic decomposition formulas). If $n\geqq 1$ and $k\notin H$, then

$$ (1)\qquad {}_Hp_{ij}^{(n)}={}_{k,H}p_{ij}^{(n)}+\sum_{\nu=1}^{n-1}{}_{k,H}p_{ik}^{(\nu)}\;{}_Hp_{kj}^{(n-\nu)}; $$

$$ (2)\qquad {}_Hp_{ij}^{(n)}={}_{k,H}p_{ij}^{(n)}+\sum_{\nu=1}^{n-1}{}_Hp_{ik}^{(\nu)}\;{}_{k,H}p_{kj}^{(n-\nu)}. $$

Proof. The first term on the right side of $\left\{\begin{matrix}(1)\\(2)\end{matrix}\right\}$ represents the probability that the state k is never entered during the transition from i to j in n steps, under the taboo set H. The general term in the sum represents the probability that k is $\left\{\begin{matrix}\text{first}\\\text{last}\end{matrix}\right\}$ entered at the ν-th step of the transition. This proves the formulas. It should be stressed that by definition the taboo in any probability is *not* in force at either end of the transition so that in the above k may well be i or j. ▯

Thus the formulas (1) and (2) may be called respectively *the first entrance decomposition* and *the last exit decomposition*. The former is just another application of the method of first entrance in §4.

Corollary. If $k \notin H$, then

(3) $${}_Hp^*_{ij} = {}_{k,H}p^*_{ij} + {}_{k,H}p^*_{ik} \cdot {}_Hp^*_{kj};$$

(4) $${}_Hp^*_{ij} = {}_{k,H}p^*_{ij} + {}_Hp^*_{ik} \cdot {}_{k,H}p^*_{kj}.$$

Proof. Summing (1) and (2) from $n=1$ to $n=N$, we obtain

(5) $$\sum_{n=1}^{N} {}_Hp^{(n)}_{ij} = \sum_{n=1}^{N} {}_{k,H}p^{(n)}_{ij} + \sum_{\nu=1}^{N-1} {}_{k,H}p^{(\nu)}_{ik} \sum_{n=1}^{N-\nu} {}_Hp^{(n)}_{kj};$$

(6) $$\sum_{n=1}^{N} {}_Hp^{(n)}_{ij} = \sum_{n=1}^{N} {}_{k,H}p^{(n)}_{ij} + \sum_{\nu=1}^{N-1} {}_Hp^{(\nu)}_{ik} \sum_{n=1}^{N-\nu} {}_{k,H}p^{(n)}_{kj}.$$

Letting $N \to \infty$ we obtain (3) and (4). Whether the series converge or diverge the results are valid owing to the conventions regarding ∞. □

The consequences of these formulas are numerous and we can only deduce a few below, leaving others unspoken of. First of all, the following two propositions will usually determine the positiveness and finiteness of the quantities we deal with.

Theorem 2. The three quantities

$${}_Hp^*_{ij}, \qquad {}_{j,H}p^*_{ij}, \qquad {}_{i,H}p^*_{ij}$$

are either all positive (possibly ∞) or all zero.

Proof. If $j \notin H$, taking $k=j$ in (3) we have

(7) $${}_Hp^*_{ij} = {}_{j,H}p^*_{ij}(1 + {}_Hp^*_{jj}).$$

If $j \in H$ then ${}_Hp^*_{ij} = {}_{j,H}p^*_{ij}$. Hence in either case if one is zero so is the other. Similarly if $i \notin H$, taking $k=i$ in (4) we have

(8) $${}_Hp^*_{ij} = {}_{i,H}p^*_{ij}(1 + {}_Hp^*_{ii}).$$

The rest follows as before. □

We write $j \rightsquigarrow H$ iff there is a k in H such that $j \rightsquigarrow k$.

Theorem 3. If either $j \in H$ or $j \rightsquigarrow H$ then ${}_Hp^*_{ij} < \infty$.

Proof. If $j \in H$ then ${}_Hp^{(n)}_{ij} \leqq {}_jp^{(n)}_{ij} = f^{(n)}_{ij}$ Hence ${}_Hp^*_{ij} \leqq f^*_{ij} \leqq 1$. Next, let $k \in H$, $k \neq j$ and $j \rightsquigarrow k$ so that ${}_kp^{(m)}_{jk} = f^{(m)}_{jk} > 0$ for some $m \geqq 1$. Clearly we have

$${}_kp^{(n)}_{ij}\, p^{(m)}_{jk} \leqq {}_kp^{(n+m)}_{ik} = f^{(n+m)}_{ik}.$$

It follows that

$${}_Hp^*_{ij} \leqq {}_kp^*_{ij} \leqq \frac{1}{f^{(m)}_{jk}} \sum_{n=1}^{\infty} f^{(n+m)}_{ik} \leqq \frac{1}{f^{(m)}_{jk}}. \qquad □$$

We proceed to derive two general formulas. Dividing (5) and (6) by $\sum_{\nu=1}^{N} {}_Hp^{(\nu)}_{kj}$ and $\sum_{\nu=1}^{N} {}_Hp^{(\nu)}_{ik}$ respectively and applying § 5, Lemma A, we obtain

$$(9)\qquad \lim_{N\to\infty}\frac{\sum_{n=1}^{N}{}_Hp_{ij}^{(n)}}{\sum_{n=1}^{N}{}_Hp_{kj}^{(n)}}=\lim_{N\to\infty}\frac{\sum_{n=1}^{N}{}_{k,H}p_{ij}^{(n)}}{\sum_{n=1}^{N}{}_Hp_{kj}^{(n)}}+{}_{k,H}p_{ik}^{*};$$

$$(10)\qquad \lim_{N\to\infty}\frac{\sum_{n=1}^{N}{}_Hp_{ij}^{(n)}}{\sum_{n=1}^{N}{}_Hp_{ik}^{(n)}}=\lim_{N\to\infty}\frac{\sum_{n=1}^{N}{}_{k,H}p_{ij}^{(n)}}{\sum_{n=1}^{N}{}_Hp_{ik}^{(n)}}+{}_{k,H}p_{kj}^{*};$$

provided that $k\notin H$ and the limit on the right side exists. These two formulas are completely general, from which simpler ones will emerge under certain mild assumptions, as we shall now see.

We write $i\overset{H}{\rightsquigarrow}j$ in case ${}_Hp_{ij}^{*}>0$; we write $i\overset{H}{\leftrightsquigarrow}j$ in case $i\overset{H}{\rightsquigarrow}j$ and $j\overset{H}{\rightsquigarrow}i$ and we say then that *i and j communicate under the taboo H*.

Theorem 4. If $i\notin H, j\notin H$ and $i\overset{H}{\rightsquigarrow}j$ then

$$(11)\qquad \lim_{N\to\infty}\frac{1+\sum_{n=1}^{N}{}_Hp_{jj}^{(n)}}{1+\sum_{n=1}^{N}{}_Hp_{ii}^{(n)}}=\frac{{}_{i,H}p_{ij}^{*}}{{}_{j,H}p_{ij}^{*}}=\lim_{N\to\infty}\frac{1+\sum_{n=1}^{N}{}_{i,H}p_{jj}^{(n)}}{1+\sum_{n=1}^{N}{}_{j,H}p_{ii}^{(n)}}.$$

Proof. Since ${}_Hp_{ij}^{*}>0$ by hypothesis we have $0<{}_{j,H}p_{ij}^{*}<\infty$ and ${}_{i,H}p_{ij}^{*}>0$ by Theorems 2 and 3. Putting $k=i$ in (10) we have

$$\lim_{N\to\infty}\frac{\sum_{n=1}^{N}{}_Hp_{ij}^{(n)}}{\sum_{n=1}^{N}{}_Hp_{ii}^{(n)}}=\frac{{}_{i,H}p_{ij}^{*}}{{}_Hp_{ii}^{*}}+{}_{i,H}p_{ij}^{*};$$

consequently

$$(12)\qquad \lim_{N\to\infty}\frac{\sum_{n=1}^{N}{}_Hp_{ij}^{(n)}}{1+\sum_{n=1}^{N}{}_Hp_{ii}^{(n)}}=\lim_{N\to\infty}\frac{\sum_{n=1}^{N}{}_Hp_{ii}^{(n)}}{\sum_{n=0}^{N}{}_Hp_{ii}^{(n)}}\left(\frac{1}{{}_Hp_{ii}^{*}}+1\right){}_{i,H}p_{ij}^{*}={}_{i,H}p_{ij}^{*}.$$

For if ${}_Hp_{ii}^{*}=\infty$ then the limit on the right side is 1 while if ${}_Hp_{ii}^{*}<\infty$ it cancels out the factor in brackets. Similarly putting $k=j$ in (9) we obtain

$$(13)\qquad \lim_{N\to\infty}\frac{\sum_{n=1}^{N}{}_Hp_{ij}^{(n)}}{1+\sum_{n=1}^{N}{}_Hp_{jj}^{(n)}}={}_{j,H}p_{ij}^{*}.$$

Taking the ratio of the last two relations we have proved the first equation in (11).

To prove the second equation in (11) we may suppose that $i \neq j$. Taking $k=j$ and replacing H by the union of H and i in (5) we have

$$\sum_{n=1}^{N} {}_{i,H}p_{ij}^{(n)} = \sum_{n=0}^{N-1} {}_{i,H}p_{jj}^{(n)} \sum_{\nu=1}^{N-n} {}_{i,H}f_{ij}^{(\nu)}.$$

Hence an application of § 5, Lemma A yields

$$\lim_{N\to\infty} \frac{\sum\limits_{n=1}^{N} {}_{i,H}p_{ij}^{(n)}}{1+\sum\limits_{n=1}^{N} {}_{i,H}p_{jj}^{(n)}} = {}_{i,H}f_{ij}^{*}.$$

Similarly, taking $k=i$ and replacing H by the union of H and j in (6) we obtain

$$\lim_{N\to\infty} \frac{\sum\limits_{n=1}^{N} {}_{j,H}p_{ij}^{(n)}}{1+\sum\limits_{n=1}^{N} {}_{j,H}p_{ii}^{(n)}} = {}_{i,H}f_{ij}^{*}. \tag{14}$$

Since ${}_{H}p_{ij}^{*} > 0$ we have firstly ${}_{H}f_{ij}^{*} > 0$ and secondly ${}_{i,H}f_{ij}^{*} > 0$ by Theorem 2. Hence we may take the ratio of the last two relations and obtain the second equation in (11).

Note that the series in (14) actually converge by Theorem 3 so that we may write (14) as

$${}_{H}f_{ij}^{*} = {}_{i,H}f_{ij}^{*}\,(1+{}_{j,H}p_{ii}^{*}).$$

This will be used in Corollary 1 below. ▯

Corollary 1. If $i \notin H$, $j \notin H$ and $i \overset{H}{\rightsquigarrow} j$ then

$$\lim_{N\to\infty} \frac{1+\sum\limits_{n=1}^{N} {}_{H}p_{jj}^{(n)}}{1+\sum\limits_{n=1}^{N} {}_{H}p_{ii}^{(n)}} = \frac{{}_{i,H}p_{ij}^{*}}{{}_{j,H}p_{ij}^{*}} = \frac{{}_{i,H}p_{ji}^{*}}{{}_{j,H}p_{ji}^{*}} = \frac{1+{}_{i,H}p_{jj}^{*}}{1+{}_{j,H}p_{ii}^{*}} = \frac{{}_{H}f_{ji}^{*}\;{}_{i,H}f_{ij}^{*}}{{}_{H}f_{ij}^{*}\;{}_{j,H}f_{ji}^{*}}, \tag{15}$$

where the limit is positive and finite. Moreover, if any two of the four states i, j, k and l communicate under the taboo H and none belongs to H, then

$$\lim_{N\to\infty} \frac{\sum\limits_{n=1}^{N} {}_{H}p_{ij}^{(n)}}{\sum\limits_{n=1}^{N} {}_{H}p_{kl}^{(n)}} \tag{16}$$

exists and is positive and finite.

The second assertion follows from the first by virtue of (12) or (13). Since we can apply either (12) or (13) to the numerator as well as the denominator of the quotient in (16) and since there are four expressions for the limit (15) we can express the limit in (16) in at least 16 generally distinct ways.

Owing to Theorem 3, the numerator and denominator will converge separately with a reasonable choice of H so that the limit will simply be ${}_Hp^*_{ij}/{}_Hp^*_{kl}$; the case where H is empty is however a notable exception.

Corollary 2. Within a class

$$\lim_{N\to\infty}\frac{\sum_{n=1}^{N}p_{ij}^{(n)}}{\sum_{n=1}^{N}p_{kl}^{(n)}}$$

exists for every i, j, k and l and it is positive and finite.

According to the Corollary to Theorem 6.4 the above limit is equal to π_{ij}/π_{kl}, provided that the last quotient does not reduce to 0/0, even without the restriction to a class. The indeterminate case occurs whenever j and l are nonrecurrent or null-recurrent. In the former case the limit is p^*_{ij}/p^*_{kl} by the Corollary to Theorem 5.5. Thus the most interesting case is where j and l are null-recurrent, in particular within such a class. Corollary 2 then provides an information which the main limit theorem does not.

In a recurrent class the formulas simplify.

Theorem 5. Within a recurrent class we have

$$(17)\qquad \lim_{N\to\infty}\frac{\sum_{n=0}^{N}p_{ij}^{(n)}}{\sum_{n=0}^{N}p_{kl}^{(n)}}=\lim_{N\to\infty}\frac{\sum_{n=0}^{N}p_{jj}^{(n)}}{\sum_{n=0}^{N}p_{ll}^{(n)}}=\frac{{}_hp^*_{hj}}{{}_hp^*_{hl}}=\frac{1+{}_lp^*_{jj}}{1+{}_jp^*_{ll}}=\frac{{}_lf^*_{lj}}{{}_jf^*_{jl}}.$$

Proof. Taking H to be empty in (13) we have

$$\lim_{N\to\infty}\frac{\sum_{n=1}^{N}p_{ij}^{(n)}}{\sum_{n=0}^{N}p_{jj}^{(n)}}={}_jp^*_{ij}=f^*_{ij}=1.$$

This is already in Theorems 5.3 and 5.6. The first equation in (17) follows. Next, taking H to be empty in the first equation of (15), we have

$$(18)\qquad \lim_{N\to\infty}\frac{\sum_{n=0}^{N}p_{jj}^{(n)}}{\sum_{n=0}^{N}p_{hh}^{(n)}}=\frac{{}_hp^*_{hj}}{{}_jp^*_{hj}}=\frac{{}_hp^*_{hj}}{f^*_{hj}}={}_hp^*_{hj}.$$

Since $0<{}_hp^*_{hj}<\infty$ for every h and j by Theorems 2 and 3, the second equation in (17) follows upon taking a quotient. The rest is in (15). ▯

In order to exhibit the basic properties of the quantities ${}_hp^*_{hi}$, which are dual to ${}_ip^*_{hi}=f^*_{hi}$, in a better light we write

$$e_{hi}={}_hp^*_{hi},$$

for every h and i in the same class. We have $0<e_{hi}<\infty$. By (18) and the Corollary to Theorem 6.4, we have $e_{hi}=\pi_i/\pi_h$ if the class is positive. The multiplicative property persists in any recurrent class.

Corollary 1. We have in a recurrent class

$$e_{hj}=e_{hi}\,e_{ij};$$

in particular $e_{hh}=1$ and $e_{hi}\,e_{ih}=1$.

Corollary 2. In a recurrent chain with independent and stationary increments (§2), we have $e_{hi}=1$ for every h and i.

For in this case $p^{(n)}_{hh}=p^{(n)}_{ii}$ for every n and every h and i.

Anticipating a more general definition in §11, we define

$$m_{ij}=\sum_{\nu=1}^{\infty}\nu\, f^{(\nu)}_{ij}.$$

For $i=j$ this reduces to the previous m_{ii}. We call m_{ij} the *mean first entrance time from i to j* in case $f^*_{ij}=1$ (so that $\{f^{(\nu)}_{ij},\nu\geqq 1\}$ determines the probability distribution of a discrete, finite-valued random variable).

Theorem 6. If $f^*_{ij}=1$ then

$$\sum_k {}_jp^*_{ik}=m_{ij}.$$

Proof. It follows from the definition that

$$\sum_k {}_jp^{(n)}_{ik}=\boldsymbol{P}\{x_\nu(\omega)\neq j,\ 1\leqq\nu<n\,|\,x_0(\omega)=i\}=\sum_{\nu=n}^{\infty}f^{(\nu)}_{ij}.$$

Hence we have

$$\sum_k {}_jp^*_{ik}=\sum_k\sum_{n=1}^{\infty}{}_jp^{(n)}_{ik}=\sum_{n=1}^{\infty}\sum_k {}_jp^{(n)}_{ik}=\sum_{n=1}^{\infty}\sum_{\nu=n}^{\infty}f^{(\nu)}_{ij}=\sum_{\nu=1}^{\infty}\nu\, f^{(\nu)}_{ij}=m_{ij},$$

according to the well known theorem on the rearrangement of double series of nonnegative terms which has been, and will be used a good number of times. ▯

Corollary. The series $\sum_k e_{hk}$ converges in a positive class and diverges in a null class.

We now come to an extension of Theorem 7.1.

Theorem 7. Let C be a recurrent class. The only nonnegative solutions $\{u_i, i\in C\}$ of the system of equations

$$u_i = \sum_{k\in C} u_k\, p_{ki}, \qquad i\in C$$

are given by $u_i = c\, e_{hi}$ where h is any state in C and c is any nonnegative constant independent of i.

Proof. It follows from the definition that

$$\sum_{k\neq h} {}_hp_{hk}^{(n)}\, p_{ki} = {}_hp_{hi}^{(n+1)}.$$

Hence we have

$$\begin{aligned}\sum_k e_{hk}\, p_{ki} &= \sum_k \sum_{n=1}^{\infty} {}_hp_{hk}^{(n)}\, p_{ki} = \sum_{n=1}^{\infty} \sum_k {}_hp_{hk}^{(n)}\, p_{ki} \\ &= \sum_{n=1}^{\infty} \{ {}_hp_{hi}^{(n+1)} + {}_hp_{hh}^{(n)}\, p_{hi} \} = e_{hi} - {}_hp_{hi}^{(1)} + {}_hp_{hh}^{*}\, p_{hi} = e_{hi}.\end{aligned}$$

This proves that $u_i = c\, e_{hi}$ is a solution.

To prove the uniqueness let $\{u_i, i\in C\}$ be a nonnegative solution, then we have

$$u_i = \sum_{k\in C} u_k\, p_{ki}^{(n)}$$

for every $n\geqq 1$ by iterating. It follows that if one $u_i = 0$ then all $u_i = 0$; hence we may suppose that all $u_i > 0$. Now define for $n \geqq 0$:

$$q_{ij}^{(n)} = \frac{u_j}{u_i}\, p_{ji}^{(n)}.$$

Clearly we have $q_{ij}^{(n)} \geqq 0$, $\sum_j q_{ij}^{(n)} = 1$ and

$$\sum_k q_{ik}\, q_{kj}^{(n)} = \sum_k \frac{u_k}{u_i}\, p_{ki}\, \frac{u_j}{u_k}\, p_{jk}^{(n)} = \frac{u_j}{u_i} \sum_k p_{jk}^{(n)}\, p_{ki} = q_{ij}^{(n+1)}.$$

Hence be the Existence theorem (Theorem 2.4) there is a M.C. (with an arbitrary initial distribution and) with (q_{ij}) as its transition matrix. The state space is C and it is obvious that it forms one class. Since $\sum_n q_{ii}^{(n)} = \sum_n p_{ii}^{(n)} = \infty$ this class is recurrent by Theorem 5.4. We have

$$\frac{\sum_{\nu=1}^{n} q_{ih}^{(\nu)}}{\sum_{\nu=1}^{n} q_{hh}^{(\nu)}} = \frac{u_h}{u_i}\, \frac{\sum_{\nu=1}^{n} p_{hi}^{(\nu)}}{\sum_{\nu=1}^{n} p_{hh}^{(\nu)}}.$$

Applying Theorem 5 simultaneously to the old and new M.C.'s we see that as $n\to\infty$ the left side tends to 1 while the right side tends to $\frac{u_h}{u_i}\, {}_hp_{hi}^{*}$. It follows that

$$u_i = u_h\, e_{hi}.$$

For an arbitrary fixed h this proves the desired uniqueness. For two different choices of h the two solutions are proportional by Corollary 1 to Theorem 5. ▯

Thus we see that the set of numbers $\{e_{hi}, i \in C\}$ for a fixed h in C enjoys certain properties similar to those of the stationary absolute distribution $\{\pi_i, i \in C\}$ in a positive class. In fact the two sets are then proportional, as already noted. Therefore the set $\{e_{hi}\}$ may be regarded as a generalization of the stationary absolute distribution to any recurrent class. In a null class $\sum_{i \in C} e_{hi} = \infty$ so that it is not a genuine distribution. The significance of the generalization will be seen in §14.

A sequence of (finite) nonnegative numbers $u(\cdot)$ on $\boldsymbol{I}$ may be regarded as a measure as well as a function on $\boldsymbol{I}$. It will then be called *superregular* iff

$$u_i \geqq \sum_{k \in \boldsymbol{I}} u_k p_{ki}, \qquad i \in \boldsymbol{I};$$

and *regular* or *subregular* iff the "$\geqq$" above is replaced by "$=$" or "$\leqq$" respectively. These definitions are in obvious duality with the corresponding ones for functions given in §7. We may now summarize some of the preceding results as follows.

Theorem 8. In any class C, and for any h in C, the measure $e_h(\cdot)$ defined by $e_h(i) = e_{hi}$ is superregular. It is regular if and only if C is recurrent, and then it is the only regular measure apart from a constant factor. It is a finite measure if and only if C is positive-recurrent.

To prove the first assertion, we need only re-examine the first few lines of the preceding proof. The rest has already been proved above.

An important property of $e_h(\cdot)$ in a class C is as follows. For any subset H of C and every h and j in H we have

$$(19) \qquad e_{hj} = \sum_{i \in H} e_{hi}\, {}_H p_{ij}^{*}.$$

This follows from the following formula obtained by considering the last exit from H:

$$(20) \qquad {}_h p_{hj}^{(n)} = {}_H p_{hj}^{(n)} + \sum_{v=1}^{n} \sum_{i \in H - \{h\}} {}_h p_{hi}^{(v)}\, {}_H p_{ij}^{(n-v)}.$$

Notes. The notion of taboo states is of course implicit in a random walk with absorbing barriers (see §5). To enter a certain state without having entered a taboo state is to convert the latter into an absorbing barrier and consider the transition in the modified scheme. In terms of the transition matrix, the operation is to replace a diagonal element by 1 and to annihilate all the other elements in the same row or cloumn. T. E. Harris [1] used the taboo probability ${}_i f_{ij}^{*}$ but mostly in a positive class. A systematic study with the present terminology and (slightly different) notation was given in Chung [2]. It may be submitted that the organized notation introduced for this purpose is quite essential, without which much of the algebra would not have been conceived. In the paper just cited, however, only one or two taboo states are considered and formulas are derived in an apparently *ad hoc* manner. It turns out that a general treatment basing all on the two decomposition formulas lends to greater unity and clarity. Thus all the results in the last cited paper are now generalized to allow for an arbitrary taboo set and without necessarily being confined within one class. This is done at no additional cost and in a more direct and inevitable way. The reason why many results in this section as well as in §11 hold with the taboo set H is because the matrices $({}_Hp_{ij}^{(n)})$, $i, j \in \boldsymbol{I} - H$, satisfy the "semigroup property":

$$ {}_Hp_{ij}^{(n+m)} = \sum_{k \in \boldsymbol{I}-H} {}_Hp_{ik}^{(n)} \, {}_Hp_{kj}^{(m)} $$

although the row sums may be less than one. Now many properties of the transition probabilities $(p_{ij}^{(n)}) = (p_{ij})^n$, $n \geqq 1$, do not depend on the fact that $\sum_j p_{ij} = 1$ and are therefore valied for the semigroup generated by the powers of a *substochastic* matrix $(\tilde{p}_{ij})$ with $\tilde{p}_{ij} \geqq 0$, $\sum_j \tilde{p}_{ij} \leqq 1$.

Corollary 2 to Theorem 4, the *ratio limit theorem,* is due to DOEBLIN [4], who gave a "geometrical" proof without expressing the limit. The first evaluation of the limit was given in CHUNG [1]. Theorem 7 is due to DERMAN [1]; for relevant examples concerning a nonrecurrent class see DERMAN [2]; for conditions of solvability of the determining system in such a class see HARRIS [2] and VEECH [1]. Corollary 2 to Theorem 5 has a striking interpretation for recurrent random walks, for it states that the expected number of visits to any point, no matter how far or how near, before returning to the starting point is always exactly one. We recall that recurrence in one dimension is guaranteed if the expected length of the random step exists and equals zero.

The transition probabilities $q_{ij}^{(n)}$ introduced in the proof of Theorem 7 are sometimes called the "*inverse probabilities*" for the M.C., see e.g. FELLER [3, p. 373]. For a chain with independent and stationary increments we have $q_{ji}^{(n)} = p_{ij}^{(n)}$ and this reversibility is of basic importance in this case.

For further results on taboo probabilities, their generating functions and related questions see FRANK [1], VERE-JONES [1], OREY [1], PORT [1], FELLER [6].

§ 10. The generating function

Let $\{a_n, n \geqq 0\}$ be a sequence of real numbers. Its *generating function* is the power series

$$ A(u) = \sum_{n=0}^{\infty} a_n u^n $$

provided that it has a nonvanishing radius of convergence. In particular if the a_n are probabilities then the radius of convergence is at least equal to one. We shall consider u as a real variable.

If $\{a_n\}$ and $\{b_n\}$ have the generating functions $A(u)$ and $B(u)$ respectively, then the sequence $\{c_n\}$ defined by

$$c_n = \sum_{k=0}^{n} a_k\, b_{n-k}, \qquad n \geqq 0$$

has the generating function $C(u) = A(u)\, B(u)$. The sequence $\{c_n\}$ is called the *convolution* of the two sequences $\{a_n\}$ and $\{b_n\}$. The virtue of the generating function lies therefore in converting the operation of convolution into that of ordinary multiplication. The following theorems are special cases of well known Abelian and Tauberian theorems.

Theorem 1. Suppose that $a_n \geqq 0$. Then we have

$$\sum_{n=0}^{\infty} a_n = \lim_{u \uparrow 1} A(u)$$

in the sense that either the two members are finite and equal or both are ∞.

Theorem 2. Suppose that $a_n \geqq 0$. Then

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} a_n = \lambda$$

if and only if

$$\lim_{u \uparrow 1} (1-u)\, A(u) = \lambda$$

where $\lambda \geqq 0$.

We now introduce the generating function

$$_H\hat{p}_{ij}(u) = \sum_{n=0}^{\infty} {_H p_{ij}^{(n)}}\, u^n, \qquad |u| < 1.$$

Equation (5.4) then becomes

$$\hat{p}_{ij} - \delta_{ij} = \hat{f}_{ij}\, \hat{p}_{jj}.$$

In particular if $i = j$ we have

$$\hat{p}_{ii} = \frac{1}{1 - \hat{f}_{ii}}. \tag{1}$$

Applying Theorem 1 to the last equation, we obtain

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \lim_{u \uparrow 1} \frac{1}{1 - \hat{f}_{ii}(u)} = \frac{1}{1 - \hat{f}_{ii}(1)} = \frac{1}{1 - f_{ii}^{*}}.$$

Thus we have a new proof of Theorem 5.4. Now assume that $f_{ii}^{*}=1$; applying Theorem 2 and using l'Hospital's rule, we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}p_{ij}^{(n)}=\lim_{u\uparrow 1}\frac{1-u}{1-\hat{f}_{ii}(u)}=\lim_{u\uparrow 1}\frac{-1}{-\frac{d}{du}\hat{f}_{ii}(u)}=\frac{1}{\sum\limits_{n=1}^{\infty}nf_{ii}^{(n)}}=\frac{1}{m_{ii}}.$$

If $i\neq j$ we have

$$\hat{p}_{ij}=\hat{f}_{ij}\hat{p}_{jj} \tag{2}$$

hence it follows that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}p_{ij}^{(n)}=\lim_{u\uparrow 1}(1-u)\hat{p}_{ij}(u)=\hat{f}_{ij}(1)\lim_{u\uparrow 1}(1-u)\hat{p}_{jj}(u)=\frac{f_{ij}^{*}}{m_{jj}}.$$

This is the Corollary to Theorem 6.4. In this light the main limit theorem of §6 is therefore a sharpening of the Tauberian theorem for the case in question.

We now try an application to the ratio limit theorem. Let $i\neq j$; taking H to be empty and replacing, j, k by i, j in (9.1) we have

$$p_{ii}^{(n)}={}_{j}p_{ii}^{(n)}+\sum_{\nu=1}^{n-1}f_{ij}^{(\nu)}p_{ji}^{(n-\nu)}.$$

Passing to generating functions we have

$$\hat{p}_{ii}={}_{j}\hat{p}_{ii}+\hat{f}_{ij}\hat{p}_{ji}.$$

Substituting from (2),

$$\hat{p}_{ii}={}_{j}\hat{p}_{ii}+\hat{f}_{ij}\hat{f}_{ji}\hat{p}_{ii}$$

or

$$\hat{p}_{ii}=\frac{{}_{j}\hat{p}_{ii}}{1-\hat{f}_{ij}\hat{f}_{ji}}. \tag{3}$$

The denominator being symmetric in i and j, upon interchanging i and j in the above equation and taking the quotient, we obtain

$$\frac{\hat{p}_{jj}}{\hat{p}_{ii}}=\frac{{}_{i}\hat{p}_{jj}}{{}_{j}\hat{p}_{ii}}.$$

If $i\leftrightsquigarrow j$ we know from Theorem 9.3 that ${}_{j}\hat{p}_{ii}(1)=1+{}_{j}p_{ii}^{*}<\infty$, ${}_{i}\hat{p}_{jj}(1)=1+{}_{i}p_{jj}^{*}<\infty$; hence

$$\lim_{u\uparrow 1}\frac{\hat{p}_{jj}(u)}{\hat{p}_{ii}(u)}=\frac{1+{}_{i}p_{jj}^{*}}{1+{}_{j}p_{ii}^{*}}.$$

It remains to prove that

$$\lim_{N\to\infty}\frac{\sum\limits_{n=0}^{N}p_{jj}^{(n)}}{\sum\limits_{n=0}^{N}p_{ii}^{(n)}}=\lim_{u\uparrow 1}\frac{\hat{p}_{jj}(u)}{\hat{p}_{ii}(u)}$$

to obtain a part of Corollary 1 to Theorem 9.4, which in turn certifies the truth of the preceding relation. But it is in general *not* true that

$$\lim_{N\to\infty}\frac{\sum\limits_{n=0}^{N} b_n}{\sum\limits_{n=0}^{N} a_n}=\lim_{u\uparrow 1}\frac{B(u)}{A(u)}.$$

The most general Tauberian theorem known of this kind, due to KARAMATA, requires the following condition on A:

$$A(u)=\frac{1}{(1-u)^{\sigma}}\,L\left(\frac{1}{1-u}\right),\qquad \sigma\geqq 0;$$

where

$$\lim_{x\to\infty}\frac{L(c\,x)}{L(x)}=1$$

for every $c>0$ (see HARDY [1]).

To avoid a direct application of this restricted Tauberian theorem we may proceed as follows. Let

$$\left.\begin{aligned} A(u)&=\hat{f}_{ij}(u)\,\hat{f}_{ji}(u)=\sum_{n=1}^{\infty} a_n\,u^n,\\ B(u)&=\frac{1}{1-A(u)}=\sum_{n=0}^{\infty} b_n\,u^n,\end{aligned}\right\}\qquad |u|<1.$$

Assume that i and j belong to the same recurrent class so that $\lim\limits_{u\uparrow 1} A(u)=f_{ij}^{*}\,f_{ji}^{*}=1$, then $\lim\limits_{u\uparrow 1} B(u)=\infty$ and it follows from Theorem 1 that $\sum\limits_{n=0}^{\infty} b_n=\infty$. Furthermore we have

$$b_n=\sum_{\nu=1}^{n} b_{n-\nu}\,a_\nu,\qquad n\geqq 1.$$

Since $\sum\limits_{n=1}^{\infty} a_n=1$ it follows by induction on n that $0\leqq b_n\leqq 1$. Writing furthermore

$${}_H\hat{P}_{ij}(u)=\frac{1}{1-u}\,{}_H\hat{p}_{ij}(u)=\sum_{\nu=0}^{\infty}\left(\sum_{n=0}^{\nu}{}_Hp_{ij}^{(n)}\right)u^{\nu}$$

the equation (3) may be written as

$$\hat{P}_{ii}={}_j\hat{P}_{ii}\,B.$$

Now we return to convolution of sequences, namely if we equate coefficients in the last equation we obtain

$$\sum_{n=0}^{N} p_{ii}^{(n)}=\sum_{\nu=0}^{N}\left(\sum_{n=0}^{\nu}{}_jp_{ii}^{(n)}\right)b_{N-\nu}. \tag{4}$$

Consequently we have

$$(5)\qquad \frac{\sum\limits_{n=0}^{N} p_{jj}^{(n)}}{\sum\limits_{n=0}^{N} p_{ii}^{(n)}} = \frac{\sum\limits_{\nu=0}^{N} \left(\sum\limits_{u=0}^{N-\nu} {}_ip_{jj}^{(n)}\right) b_\nu}{\sum\limits_{\nu=0}^{N} \left(\sum\limits_{n=0}^{N-\nu} {}_jp_{ii}^{(n)}\right) b_\nu}.$$

The facts that $0 \leqq b_n \leqq 1$, $\sum\limits_{\nu=0}^{\infty} b_\nu = \infty$ and $\sum\limits_{n=0}^{\infty} {}_ip_{jj}^{(n)} = 1 + {}_ip_{jj}^{*} < \infty$, $\sum\limits_{n=0}^{\infty} {}_jp_{ii}^{(n)} = 1 + {}_jp_{ii}^{*} < \infty$ are sufficient to establish that the right member in (5) tends to the limit

$$\frac{1 + {}_ip_{jj}^{*}}{1 + {}_jp_{ii}^{*}}$$

as $N \to \infty$; so does the left member and the desired result is proved.

This new proof of a previous result illustrates another use of the generating function, namely as a convenient means of deriving algebraic relations, here the formula (4), in a symbolic way. Its previous use, depending on the Abelian and Tauberian theorems, is more genuinely analytic. We conclude this discussion with yet another application in a different direction, in which the use of generating function is not indispensable but most expedient in effecting computations.

In view of Theorem 9.4 and its corollaries, and in view of the main limit theorem of §6, it is natural to inquire if the limit of an *individual* ratio like $p_{ij}^{(n)}/p_{kl}^{(n)}$ may not exist as $n \to \infty$, at least within a recurrent class of period 1. It is known that such is indeed the case in a chain with independent and stationary increments, and in certain other special types of M.C. The answer however is negative in the general case, as we now proceed to show.

Consider a M.C. with state space consisting of the element 0 and all ordered pairs (i, j) of positive integers such that $j \leqq n_i - 1$, where $n_i = 2^{i^2}$. Let $p_0 = 1$;

$$p_{00} = 2^{-1}, \qquad p_{0,(i,1)} = 2^{-i-1}, \qquad i \geqq 1;$$

$$p_{(i,j),(i,j+1)} = 1, \qquad 1 \leqq j \leqq n_i - 2; \qquad p_{(i,n_i-1),0} = 1.$$

Obviously all the states form a class of period 1. It is clear that

$$f_{00}^{(n)} = \begin{cases} 2^{-i-1} & \text{if } n = n_i, \quad i \geqq 0; \\ 0 & \text{otherwise.} \end{cases}$$

Thus $\sum\limits_{n} f_{00}^{(n)} = 1$ and the class is recurrent. Denoting the state (1, 1) by 1, we have obviously

$$p_{10}^{(n)} = p_{10}\, p_{00}^{(n-1)} = p_{00}^{(n-1)}$$

$$p_{11}^{(n)} = p_{10}\, p_{00}^{(n-2)}\, p_{01} = \tfrac{1}{4} p_{00}^{(n-2)}$$

Now if $p^{(n)}_{10}/p^{(n)}_{00}$ has a limit as $n \to \infty$ this limit must be 1, and if $p^{(n)}_{11}/p^{(n)}_{00}$ has a limit it must be positive and finite by Theorem 9.5. Hence by the equations above the ratios $p^{(n)}_{00}/p^{(n-1)}_{00}$ and $p^{(n)}_{00}/p^{(n-2)}_{00}$ must have finite limits. We have thus reduced the problem to the existence of the last mentioned limits.

To show that neither of these limits exists, finite or infinite, we use generating functions and the equation (1) above. We have, if $i \geqq 3$ and $n = n_i$, $n_i - 1$ or $n_i - 2$:

$$\begin{aligned} p^{(n)}_{00} &= \text{coefficient of } u^n \text{ in } \frac{1}{1-\hat{f}_{00}(u)} \\ &= f^{(n)}_{00} + \text{coefficient of } u^n \text{ in } \frac{1}{1-\sum\limits_{\nu=1}^{n-1} f^{(\nu)}_{00} u^\nu} \\ &= f^{(n)}_{00} + \text{coefficient of } u^n \text{ in } \frac{1}{1-\sum\limits_{\nu=1}^{n_{i-1}} f^{(\nu)}_{00} u^\nu} \\ &= f^{(n)}_{00} + \text{coefficient of } u^n \text{ in } \sum_{s=2^{2i-1}}^{\infty} \left\{ \sum_{\nu=1}^{n_{i-1}} f^{(\nu)}_{00}\, u^\nu \right\}^s . \end{aligned}$$

For the last step we note that $n_i/n_{i-1} = 2^{2i-1}$. Hence we have

$$\begin{aligned} 0 \leqq p^{(n)}_{00} - f^{(n)}_{00} \leqq \sum_{s=2^{2i-1}}^{\infty} \left\{ \sum_{\nu=1}^{n_{i-1}} f^{(\nu)}_{00} \right\}^s &= \sum_{s=2^{2i-1}}^{\infty} (1-2^{-i})^s \\ &= \frac{(1-2^{-i})^{2^{2i-1}}}{2^{-i}} < 2^i \left(e^{-2^{-i}}\right)^{2^{2i-1}} = 2^i e^{-2^{i-1}}. \end{aligned}$$

It follow that

$$\frac{p^{(n_i)}_{00}}{p^{(n_i-\delta)}_{00}} \geqq \frac{2^{-i-1}}{2^i e^{-2^{i-1}}} = 2^{-2i-1} e^{2^{i-1}}$$

for $\delta = 1$ and $\delta = 2$. Therefore we conclude that

$$\overline{\lim_{n\to\infty}} \frac{p^{(n)}_{00}}{p^{(n-\delta)}_{00}} = \infty$$

and since $p^{(n)}_{00} \leqq 1$ the corresponding limit cannot exist, finite or infinite.

The following sufficient condition for the individual ratio limit theorem is illuminating and furnishes another example of the solidarity among the states in a recurrent class.

Theorem 3. Suppose that in a recurrent class we have for a certain state 0:

$$\lim_{n\to\infty} p^{(n+1)}_{00}/p^{(n)}_{00} = 1, \tag{6}$$

then for any states i, j, k, l in the class we have

$$\lim_{n\to\infty} p^{(n)}_{ij}/p^{(n)}_{kl} = {}_l p^*_{ij} = e_{lj}. \tag{7}$$

Proof. Let us put

(8) $$e_{ij}^{(n)} = {}_ip_{ij}^{(n)}$$

so that in our previous notation we have $e_{ij} = {}_ip_{ij}^* = \sum_{n=1}^{\infty} e_{ij}^{(n)}$. We have for each j:

(9) $$p_{0j}^{(n)} = \sum_{\nu=1}^{n} p_{00}^{(n-\nu)} e_{0j}^{(\nu)}$$

as the "dual" to (5.4). After division by $p_{00}^{(n)}$, it follows by (6) and FATOU's lemma that

(10) $$\varliminf_{n\to\infty} p_{0j}^{(n)}/p_{00}^{(n)} \geqq e_{0j}.$$

Now consider the equation

$$p_{00}^{(n+m)}/p_{00}^{(n)} = \sum_{k} \{p_{0k}^{(n)}/p_{00}^{(n)}\} p_{k0}^{(m)}.$$

It follows by (10) that

$$1 \geqq \varlimsup_{n\to\infty} \{p_{0j}^{(n)}/p_{00}^{(n)}\} p_{j0}^{(m)} + \sum_{k \neq j} e_{0k}\, p_{k0}^{(m)}.$$

The last written sum without the restriction "$k \neq j$" is equal to $e_{00} = f_{00}^* = 1$ by Theorem 9.3. Choosing m so that $p_{j0}^{(m)} > 0$, it follows that the upper limit appearing above does not exceed e_{0j} and so in combination with (10) we have

(11) $$\lim_{n\to\infty} p_{0j}^{(n)}/p_{00}^{(n)} = e_{0j}.$$

Similarly we obtain, by arguments which are dual to the above:

(12) $$\lim_{n\to\infty} p_{j0}^{(n)}/p_{00}^{(n)} = f_{j0}^* = 1.$$

Furthermore it follows from the equation

$$p_{jj}^{(n)} = \sum_{\nu=1}^{n-1} f_{j0}^{(\nu)}\, p_{0j}^{(n-\nu)} + {}_0p_{jj}^{(n)},$$

and by (11) that

(13) $$\varliminf_{n\to\infty} \{p_{jj}^{(n)}/p_{00}^{(n)}\} \geqq f_{j0}^*\, e_{0j} = e_{0j}.$$

On the other hand, by applying (11) and (13) to the equation

$$p_{0j}^{(n)} = \sum_{\nu=1}^{n} f_{0j}^{(\nu)}\, p_{jj}^{(n-\nu)},$$

we obtain

$$e_{0j} \geqq f_{0j}^{(m)} \varlimsup_{n\to\infty} \{p_{jj}^{(n-m)}/p_{00}^{(n)}\} + (1 - f_{0j}^{(m)})\, e_{0j}.$$

Choosing m so that $f_{0j}^{(m)} > 0$, it follows that the upper limit appearing above, which is the same as $\varlimsup_{n\to\infty} \{p_{jj}^{(n)}/p_{00}^{(n)}\}$ by (6), does not exceed e_{0j} and so in combination with (13) we obtain

$$\lim_{n\to\infty}\{p_{jj}^{(n)}/p_{00}^{(n)}\}=e_{0j}. \tag{14}$$

Consequently, the condition (6) is now valid with 0 replaced by an arbitrary state, and so are the relations (11), (12) and (14) under the same replacement. The existence of the limit in (7) is thereby established, and its value of is course the same in as (9.17), evaluated with $h=l$.

Notes. Theorem 1 is elementary; Theorem 2 is Theorem 96 in Hardy [1]; the other Tauberian theorem cited is a consequence of Theorem 108 there.

It may be mentioned that a somewhat more elaborate use of the generating function involving Wiener's Theorem on absolutely convergent Fourier series leads to a proof of the main limit theorem (Theorem 6.1) in the case $m_{ii}<\infty$. Similar methods have been used by Feller [2] in obtaining further results mentioned in the notes on §6.

The new proof of the ratio limit theorem is the original one announced in Chung [1].

For the individual ratio limit theorem in a chain with independent and stationary increments, stated in a more specific form but equivalent to the general result, see Chung and Erdös [1]; there is a complement due to Kesten [1]. For the other special case see Karlin and McGregor [1]. The general question was raised in Chung [2]. The simple reduction to the more tractable problem concerning a single state is due to the author; the actual example is due to F. J. Dyson (private communication). Theorem 3 is due to Orey [2]; see also Kingman and Orey [1].

For results concerning the application of the Karamata Tauberian theorem see Lamperti [1].

§ 11. The moments of first entrance time distributions

If $f_{ij}^{*}=1$, then the sequence $\{f_{ij}^{(n)}, n\geqq 1\}$ determines a discrete probability distribution called the *first entrance time distribution from i to j*. (For $i=j$ this has also already been called the recurrence time distribution of i in §6.) Thus for each p, $\sum_{n=1}^{\infty} n^p f_{ij}^{(n)}$ is the moment of order p of this distribution; for $p=1$ this is the m_{ij} defined in §9. More generally, let H be the taboo set; we write

$${}_Hm_{ij}^{(p)}=\sum_{n=1}^{\infty} n^p\, {}_Hf_{ij}^{(n)}.$$

If $i\overset{H}{\rightsquigarrow}j$ then ${}_Hm_{ij}^{(p)}/{}_Hf_{ij}^{*}$ is the moment of order p of the *conditional first entrance time distribution* determined by $\{{}_Hf_{ij}^{(n)}/{}_Hf_{ij}^{*}, n\geqq 1\}$.

Theorem 1. Let J be an arbitrary nonempty set of states and H be an arbitrary, possibly empty, set of states that is disjoint from J and such that any two states in J communicate under the taboo H. Let $p \geqq 0$ be a fixed number. If we have

$$(1) \qquad {}_H m_{ij}^{(p)} < \infty, \qquad {}_H m_{ji}^{(p)} < \infty$$

for some i and j in J, then the same is true of every i and j in J (i and j need not be distinct in either instance).

Proof. In this proof we shall write, for the sake of simplicity,

$${}_k\varphi_{ij}^{(n)} = {}_{k,H}f_{ij}^{(n)}, \qquad {}_k\varphi_{ij}^{*} = {}_{k,H}f_{ij}^{*}, \qquad {}_k\mu_{ij}^{(p)} = {}_{k,H}m_{ij}^{(p)}.$$

If $i \neq j$, $i \notin H$ we have from (9.1), replacing H, k there by $H \cup \{j\}$, i:

$$(2) \qquad \varphi_{ij}^{(n)} = {}_i\varphi_{ij}^{(n)} + \sum_{\nu=1}^{n-1} {}_j\varphi_{ii}^{(\nu)}\, \varphi_{ij}^{(n-\nu)};$$

similarly, replacing H, j, k by $H \cup \{i\}$, i, j:

$$\varphi_{ii}^{(n)} = {}_j\varphi_{ii}^{(n)} + \sum_{\nu=1}^{n-1} {}_i\varphi_{ij}^{(\nu)}\, \varphi_{ji}^{(n-\nu)}.$$

Multiplying through by n^p and summing over n, we obtain

$$(3) \qquad \mu_{ij}^{(p)} = {}_i\mu_{ij}^{(p)} + \sum_{\nu=1}^{\infty} {}_j\varphi_{ii}^{(\nu)} \sum_{s=1}^{\infty} (s+\nu)^p\, \varphi_{ij}^{(s)},$$

$$(4) \qquad \mu_{ii}^{(p)} = {}_j\mu_{ii}^{(p)} + \sum_{\nu=1}^{\infty} {}_i\varphi_{ij}^{(\nu)} \sum_{s=1}^{\infty} (s+\nu)^p\, \varphi_{ji}^{(s)}.$$

These equations are to be interpreted according to our conventions regarding ∞ (see §9). We divide the rest of the proof into three steps.

(i) We show that (1) implies that

$$\mu_{ii}^{(p)} < \infty, \qquad \mu_{jj}^{(p)} < \infty.$$

If $i = j$ there is nothing to prove. If $i \neq j$ we have from (3), since $(s+\nu)^p \geqq \nu^p$,

$$\mu_{ij}^{(p)} \geqq {}_i\mu_{ij}^{(p)} + \sum_{\nu=1}^{\infty} {}_j\varphi_{ii}^{(\nu)}\, \nu^p \sum_{s=1}^{\infty} \varphi_{ij}^{(s)} \geqq {}_i\mu_{ij}^{(p)} + \varphi_{ij}^{*}\, {}_j\mu_{ii}^{(p)}.$$

Since $\varphi_{ij}^{*} > 0$ it follows that

$$(5) \qquad {}_i\mu_{ij}^{(p)} < \infty, \qquad {}_j\mu_{ii}^{(p)} < \infty.$$

Using the inequality $(s+\nu)^p \leqq 2^p (s^p + \nu^p)$ we have

$$\mu_{ii}^{(p)} \leqq {}_j\mu_{ii}^{(p)} + 2^p \sum_{\nu=1}^{\infty} {}_i\varphi_{ij}^{(\nu)} \{\mu_{ji}^{(p)} + \nu^p\} = {}_j\mu_{ii}^{(p)} + 2^p \{{}_i\varphi_{ij}^{*}\, \mu_{ji}^{(p)} + {}_i\mu_{ij}^{(p)}\}.$$

Consequently we infer from (5) and the second assumption in (1) that $\mu_{ii}^{(p)} < \infty$. By symmetry $\mu_{jj}^{(p)} < \infty$.

(ii) We show that $\mu_{ii}^{(p)} < \infty$ implies that

$$\mu_{ki}^{(p)} < \infty, \qquad \mu_{ik}^{(p)} < \infty \tag{6}$$

for every k in J. From (4) with k instead of j we have, since $(s+\nu)^p \geqq \max(s^p, \nu^p)$,

$$\mu_{ii}^{(p)} \geqq \max\{{}_i\varphi_{ik}^* \, \mu_{ki}^{(p)}, \; \varphi_{ki}^* \, {}_i\mu_{ik}^{(p)}\}.$$

Since $\varphi_{ki}^* > 0$ and we have ${}_i\varphi_{ik}^* > 0$ by Theorem 9.2, it follows that

$$\mu_{ki}^{(p)} < \infty, \qquad {}_i\mu_{ik}^{(p)} < \infty$$

for every k in J. Next we have, writing k for j in (2) and summing over n from 1 to N:

$$\begin{aligned}\sum_{n=1}^{N} n^p \varphi_{ik}^{(n)} &= \sum_{n=1}^{N} n^p \, {}_i\varphi_{ik}^{(n)} + \sum_{\nu=1}^{N-1} {}_k\varphi_{ii}^{(\nu)} \sum_{s=1}^{N-\nu} (s+\nu)^p \varphi_{ik}^{(s)} \\ &= \sum_{n=1}^{N} n^p \, {}_i\varphi_{ik}^{(n)} + \sum_{\nu=1}^{N-1} {}_k\varphi_{ii}^{(\nu)} \sum_{s=1}^{N-\nu} s^p \varphi_{ik}^{(s)} + \sum_{s=1}^{N-1} {}_k\varphi_{ii}^{(\nu)} \sum_{s=1}^{N-\nu} \{(s+\nu)^p - s^p\} \varphi_{ik}^{(s)}.\end{aligned}$$

If $0 \leqq p \leqq 1$, then $(s+\nu)^p \leqq s^p + \nu^p$ and the above equation yields

$$\sum_{n=1}^{N} n^p \varphi_{ik}^{(n)} \left\{1 - \sum_{\nu=1}^{N-n} {}_k\varphi_{ii}^{(\nu)}\right\} \leqq {}_i\mu_{ik}^{(p)} + {}_k\mu_{ii}^{(p)} \varphi_{ik}^* < \infty, \tag{7}$$

since ${}_k\mu_{ii}^{(p)} \leqq \mu_{ii}^{(p)} < \infty$. Letting $N \to \infty$ we obtain

$$\mu_{ik}^{(p)} (1 - {}_k\varphi_{ii}^*) < \infty.$$

Since ${}_k\varphi_{ii}^*$ and ${}_i\varphi_{ik}^*$ are probabilities of disjoint events if $i \neq k$, we have $1 - {}_k\varphi_{ii}^* \geqq {}_i\varphi_{ik}^* > 0$. It follows that $\mu_{ik}^{(p)} < \infty$.

If $p > 1$, we use the inequality

$$(s+\nu)^p - s^p \leqq p\nu(s+\nu)^{p-1} \leqq p\, 2^{p-1} (\nu\, s^{p-1} + \nu^p).$$

Instead of (7) we have then

$$\sum_{n=1}^{N} n^p \varphi_{ik}^{(n)} \left\{1 - \sum_{\nu=1}^{N-n} {}_k\varphi_{ii}^{(\nu)}\right\} \leqq {}_i\mu_{ik}^{(p)} + p\, 2^{p-1} ({}_k\mu_{ii}^{(1)} \mu_{ik}^{(p-1)} + {}_k\mu_{ii}^{(p)} \varphi_{ik}^*). \tag{8}$$

We are going to use induction on the integral part $[p]$ or p. Suppose we have proved that $\mu_{ii}^{(p-1)} < \infty$ implies $\mu_{ik}^{(p-1)}$ for every k. If $\mu_{ii}^{(p)} < \infty$ then $\mu_{ii}^{(p-1)} < \infty$ and hence by the induction hypothesis $\mu_{ik}^{(p-1)} < \infty$ for every k in J. Since $p > 1$ the assumption $\mu_{ii}^{(p)} < \infty$ also implies $\mu_{ii}^{(1)} < \infty$ and *a fortiori* ${}_k\mu_{ii}^{(1)} < \infty$. Hence it follows upon letting $N \to \infty$ in (8) that $\mu_{ik}^{(p)} < \infty$. Since we have already proved that the assertion "$\mu_{ii}^{(p)} < \infty$ implies $\mu_{ik}^{(p)} < \infty$ for every k in J" is true for all p with $0 \leqq p \leqq 1$, it is now true for all $p \geqq 0$.

(iii) Since (6) is true we have by part (i) that $\mu_{kk}^{(p)} < \infty$ for every k in J. Substituting k, l for i, k in part (ii) we conclude that $\mu_{lk}^{(p)} < \infty$, $\mu_{kl}^{(p)} < \infty$ for every l in J. The theorem is therefore completely proved. □

Corollary 1. Within a class if

$$m_{ij}^{(p)} < \infty, \qquad m_{ji}^{(p)} < \infty$$

for a pair of (not necessarily distinct) states, then the same is true of any pair of (not necessarily distinct) states. Thus in particular the property that $m_{ii}^{(p)} < \infty$ for a state i is a class property.

For an integer p our formulas (3) and (4) lead to explicit relations. More general ones can be obtained from Theorem 9.1, but we content ourselves with stating the following.

Corollary 2. For integer $p \geqq 0$ and $i \neq j$,

$$ {}_H m_{ii}^{(p)} = \sum_{r=0}^{p} \binom{p}{r} {}_{i,H} m_{ij}^{(p-r)}\, {}_H m_{ji}^{(r)} + {}_{j,H} m_{ii}^{(p)} $$

$$ {}_H m_{ij}^{(p)} = \sum_{r=0}^{p} \binom{p}{r} {}_{j,H} m_{ii}^{(p-r)}\, {}_H m_{ij}^{(r)} + {}_{i,H} m_{ij}^{(p)}\,. $$

Corollary 1 permits a further classification of all classes according to the existence of moments (in a recurrent class) or conditional moments (in a nonrecurrent class). The *order* of a state i is the supremum of all $p \geqq 0$ such that $m_{ii}^{(p)} < \infty$; clearly it is also the infimum of all $p \geqq 0$ such that $m_{ii}^{(p)} = \infty$. All the states in a class have the same order which may, therefore, be called the order of the class. If the order of i is p then $m_{ii}^{(p)}$ may be finite or infinite; for example a state of order 1 may be positive or null. This ambiguity makes it inconvenient to use the order in practice.

Theorem 2. Given any $p \geqq 0$ there exists a M.C. of which the states form a class of order p. Furthermore we may stipulate that the class be recurrent or nonrecurrent; and if $p > 0$ that $m_{ii}^{(p)}$ be finite or infinite for all i.

Proof. It follows from the discussion in §8, and also easily otherwise, that there is a M.C. of which the states form one class and with arbitrarily preassigned $\{f_{00}^{(n)}, n \geqq 1\}$ for a certain state 0. Let us then assign

$$ f_{00}^{(n)} = \frac{c}{n^{p+1}[\lg(n+1)]^2}, \qquad p \geqq 0; $$

or

$$ f_{00}^{(n)} = \frac{c'}{n^{p+1}}, \qquad p > 0; $$

where c and c' are positive numbers such that $\sum_{n=1}^{\infty} f_{00}^{(n)} \leqq 1$. The state 0 and therefore the M.C. is recurrent if and only if equality holds. It is obvious that the order is p in either case; and if $p>0$, $m_{00}^{(p)}<\infty$ in the first case and $m_{00}^{(p)}=\infty$ in the second case. □

We turn to the study of some special moments without attempt at generalization. We write ${}_H m_{ij}$ for ${}_H m_{ij}^{(1)}$ for brevity.

Theorem 3. If i, j and k belong to the same positive class and $j\neq k$, we have

$$m_{ik}+m_{kj}-m_{ij}={}_k f_{ij}^{*}(m_{jk}+m_{kj}). \tag{9}$$

Proof. We have from (9.1)

$$f_{ij}^{(n)}={}_k f_{ij}^{(n)}+\sum_{\nu=1}^{n-1} {}_j f_{ik}^{(\nu)} f_{kj}^{(n-\nu)}. \tag{10}$$

Summing over n we obtain, as a special case of (9.3),

$$f_{ij}^{*}={}_k f_{ij}^{*}+{}_j f_{ik}^{*} f_{kj}^{*}.$$

Since $f_{ij}^{*}=f_{kj}^{*}=1$, this reduces to

$${}_k f_{ij}^{*}+{}_j f_{ik}^{*}=1.$$

This relation, valid within any recurrent class, has an obvious intuitive meaning. Next, multiplying (10) through by n and summing over n, we obtain

$$m_{ij}={}_k m_{ij}+{}_j m_{ik}+{}_j f_{ik}^{*} m_{kj}.$$

Interchanging j and k in the above and comparing the results, we obtain (9). □

Corollary 1.

$$m_{jj}={}_j f_{jk}^{*}(m_{jk}+m_{kj}). \tag{11}$$

Corollary 2.

$${}_k p_{ij}^{*}=\frac{{}_k f_{ij}^{*}}{{}_j f_{jk}^{*}}=\frac{m_{ik}+m_{kj}-m_{ij}}{m_{jj}}. \tag{12}$$

Proof. Specializing (9.7) and (9.8) we have

$${}_k p_{ij}^{*}={}_k f_{ij}^{*}(1+{}_k p_{jj}^{*}),$$

$$1=f_{jk}^{*}={}_j f_{jk}^{*}(1+{}_k p_{jj}^{*});$$

from which the first equation in (12) follows upon division. The second equation follows upon dividing (9) by (11). □

Putting $i=k$ in (12) we have

$$\frac{m_{kk}}{m_{jj}}=\frac{{}_k f_{kj}^{*}}{{}_j f_{jk}^{*}}={}_k p_{kj}^{*}.$$

This is a special case of (9.18), on account of the Corollary to Theorem 6.4.

Theorem 4. If i, j and k belong to a positive class of period one, then

$$\sum_{n=1}^{\infty}\{p_{ik}^{(n)}-p_{jk}^{(n)}\}=\frac{m_{jk}-m_{ik}}{m_{kk}}.$$

Proof. We have by definition

$$\sum_{l\neq i} {}_ip_{jl}^{(n)}\,p_{lk}={}_ip_{jk}^{(n+1)}.$$

Summing over n, we obtain

$$\sum_{l\neq i} {}_ip_{jl}^{*}\,p_{lk}={}_ip_{jk}^{*}-p_{jk}$$

or

$$\sum_{l} {}_ip_{jl}^{*}\,p_{lk}={}_ip_{jk}^{*}+p_{ik}-p_{jk}.$$

The series converge by Theorem 9.6, since every $m_{ji}<\infty$ in a positive class by Theorem 1. Iterating the last equation, we have

$$\sum_{l} {}_ip_{jl}^{*}\,p_{lk}^{(n)}={}_ip_{jk}^{*}+\sum_{\nu=1}^{n}\{p_{ik}^{(\nu)}-p_{jk}^{(\nu)}\}.$$

Letting $n\to\infty$, we obtain by Theorems 6.1 and 9.6 and Corollary 2 to Theorem 3:

$$\begin{aligned}\sum_{\nu=1}^{\infty}\{p_{ik}^{(\nu)}-p_{jk}^{(\nu)}\}&=\sum_{l} {}_ip_{jl}^{*}\,\frac{1}{m_{kk}}-{}_ip_{jk}^{*}\\&=\frac{m_{ji}}{m_{kk}}-\frac{m_{ji}+m_{ik}-m_{jk}}{m_{kk}}=\frac{m_{jk}-m_{ik}}{m_{kk}}.\end{aligned}$$ □

If we enumerate the states in any way with the positive integers, then *we say* $k\to\infty$ *if the integer corresponding to the state* k *tends to* ∞.

Theorem 5. If for some $p\geqq 0$ we have $m_{ij}^{(p)}<\infty$ then

$$\lim_{k\to\infty} {}_km_{ij}^{(p)}=m_{ij}^{(p)}.$$

Proof. It follows from (10) that

$$0\leqq f_{ij}^{(n)}-{}_kf_{ij}^{(n)}\leqq\sum_{\nu=1}^{n-1}p_{ik}^{(\nu)}.$$

Consequently we have for every $n\geqq 1$,

$$\lim_{k\to\infty} {}_kf_{ij}^{(n)}=f_{ij}^{(n)}.$$

We have therefore

$$\lim_{k\to\infty} {}_km_{ij}^{(p)}=\lim_{k\to\infty}\sum_{n=1}^{\infty}n^p\,{}_kf_{ij}^{(n)}=\sum_{n=1}^{\infty}n^p\,f_{ij}^{(n)}=m_{ij}^{(p)}$$

since the first series, being dominated by the second series, converges uniformly with respect to k. □

Corollary.

$$\lim_{k\to\infty} {}_j m_{ik}^{(p)} f_{kj}^* = \lim_{k\to\infty} {}_j f_{ik}^* m_{kj}^{(p)} = 0.$$

Proof. As in the proof of Theorem 1, we have

$$m_{ij}^{(p)} \geqq {}_k m_{ij}^{(p)} + \max\{{}_j f_{ik}^* m_{kj}^{(p)},\ {}_j m_{ik}^{(p)} f_{kj}^*\}.$$

Hence the corollary follows from the theorem. ▯

Theorem 6. If i belongs to a positive class C containing an infinite number of states, then

$$\lim_{\substack{j\to\infty \\ j\in C}} \frac{m_{ji}}{m_{ij}} = 0 \tag{13}$$

and

$$\sum_{j\in C} \frac{1}{m_{ij}} < \infty. \tag{14}$$

Proof. We have by (11)

$$m_{ii} = {}_i f_{ij}^* m_{ij} + {}_i f_{ij}^* m_{ji}.$$

The second term on the right side tends to 0 as $j\to\infty$, by the Corollary to Theorem 5; hence the first term tends to m_{ii} and (13) follows. Moreover $m_{jj} \leqq m_{ij} + m_{ji}$ and we have

$$\sum_{j\in C} \frac{1}{m_{jj}} = 1$$

by Theorem 7.1. From these relations and (13) follows (14). ▯

Theorem 7. If i belongs to a positive class C we have

$$m_{ii}^{(2)} = m_{ii}\left(2\sum_{j\in C} \frac{m_{ji}}{m_{jj}} - 1\right). \tag{15}$$

This is a special case of Theorem 14.7 to which we refer its proof. We draw, however, some consequences related to the preceding results.

Corollary 1. The series

$$\sum_{j\in C} \frac{m_{ji}}{m_{jj}}$$

converges for all i in C or none. The former case obtains if and only if $m_{ii}^{(2)} < \infty$.

Corollary 2. If $m_{ii}^{(2)} = \infty$ then

$$\lim_{j\to\infty} m_{ji} = \infty.$$

To illustrate the above results and also some unmentioned negative results we consider the following example which lends itself to easy calculations.

Example. Let I be the set of nonnegative integers. Let $\alpha_j > 0$, $\sum_{j=1}^{\infty}\alpha_j = 1$, $0 < \beta_j < 1$:

$$p_{0j} = \alpha_j, \quad p_{j0} = \beta_j, \quad p_{jj} = 1 - \beta_j, \quad j \geqq 1.$$

Clearly all the states form a recurrent class of period 1. The states j, $j \geqq 1$, can communicate with each other only via the state 0. Consider the system (7.1) which reduces to

$$u_0 = \sum_{j=1}^{\infty} \beta_j u_j$$

$$u_j = (1 - \beta_j) u_j + \alpha_j u_0, \quad j \geqq 1.$$

This system has an absolutely convergent solution if and only if

$$\sum_{j=1}^{\infty} \frac{\alpha_j}{\beta_j} < \infty. \tag{16}$$

Applying Theorem 7.1 we obtain all the mean recurrence times as follows:

$$m_{00} = 1 + \sum_{j=1}^{\infty} \frac{\alpha_j}{\beta_j}; \quad m_{jj} = \frac{\beta_j}{\alpha_j} m_{00}, \quad j \geqq 1.$$

In what follows let $j \geqq 1$. By a direct consideration we have $f_{j0}^{(n)} = (1 - \beta_j)^{n-1} \beta_j$; hence

$$m_{j0} = \sum_{n-1}^{\infty} n (1 - \beta_j)^{n-1} \beta_j = \frac{1}{\beta_j}.$$

It is also clear that

$${}_0f_{0j}^* = \alpha_j, \quad {}_jf_{j0}^* = \beta_j.$$

We can now compute m_{0j} by means of the equation (11):

$$m_{00} = {}_0f_{0j}^* (m_{0j} + m_{j0}).$$

Thus weobtain

$$m_{0j} = \frac{1}{\alpha_j} \left(1 + \sum_{\substack{k=1 \\ k \neq j}}^{\infty} \frac{\alpha_k}{\beta_k} \right).$$

It is clear that if $j \neq k$,

$$m_{jk} = m_{j0} + m_{0k} = \frac{1}{\beta_j} + \frac{1}{\alpha_k} \left(1 + \sum_{\substack{l=1 \\ l \neq k}}^{\infty} \frac{\alpha_l}{\beta_l} \right).$$

The conditions on $\{\alpha_j\}$ and $\{\beta_j\}$ are compatible with the sequence $\{m_{j0}\} = \{\beta_j^{-1}\}$ having any limit between 1 and ∞ (both inclusive) or no limit. A similar assertion is true of the sequence $m_{0j}\, m_{jj}^{-1}$ since

$$\frac{m_{0j}}{m_{jj}} = \frac{1}{\beta_j} - \frac{\alpha_j}{m_{00} \beta_j^2}$$

and $\lim\limits_{j\to\infty} \alpha_j \beta_j^{-1} = 0$ by (16). Next, we have

$$\frac{m_{j0}}{m_{jj}} = \frac{\alpha_j}{m_{00}\beta_j^2}\,.$$

Hence the sequence $\{m_{j0}\, m_{jj}^{-1}\}$ may have any limit between 0 and ∞ (both inclusive) or no limit. However

$$\frac{m_{j0}}{m_{0j}} = \frac{\alpha_j}{\beta_j\left(1+\sum\limits_{k\neq j}\frac{\alpha_k}{\beta_k}\right)} \leqq \frac{\alpha_j}{\beta_j}$$

and so tends to 0 as $j\to\infty$ on account of (16).

Finally, a direct consideration yields $f_{00}^{(n)} = \sum\limits_{j=1}^{\infty} \alpha_j \beta_j (1-\beta_j)^{n-2}$ for $n\geqq 2$; and consequently

$$m_{00}^{(2)} = \sum_{j=1}^{\infty} \alpha_j \left(1+\frac{1}{\beta_j}+\frac{2}{\beta_j^2}\right)$$

which checks with (15).

Theorem 4 above has the following extension which is illustrative of the method of calculation with taboo probabilities.

Theorem 8. Suppose that the M.C. is either positive-recurrent of period one or null-recurrent of any period. If $i\neq j$, $i\neq k$, then we have

$$\sum_{n=0}^{\infty} \{p_{ii}^{(n)} - p_{ji}^{(n)} + [p_{jk}^{(n)} - p_{ik}^{(n)}]\, e_{ki}\} = (\delta_{jk} + {}_iP_{jk}^{*})\, e_{ki}. \tag{17}$$

Proof. We have

$$\sum_{n=0}^{N} p_{ii}^{(n)} = \sum_{\nu=1}^{N} f_{ji}^{(\nu)} \sum_{n=0}^{N} p_{ii}^{(n)} + \sum_{\nu=N+1}^{\infty} f_{ji}^{(\nu)} \sum_{n=0}^{N} p_{ii}^{(n)};$$

$$\sum_{n=0}^{N} p_{ik}^{(n)} = \sum_{\nu=1}^{N} f_{ji}^{(\nu)} \sum_{n=0}^{N} p_{ik}^{(n)} + \sum_{\nu=N+1}^{\infty} f_{ji}^{(\nu)} \sum_{n=0}^{N} p_{ik}^{(n)};$$

$$\sum_{n=0}^{N} p_{jk}^{(n)} = \sum_{\nu=1}^{N} f_{ji}^{(\nu)} \sum_{n=0}^{N-\nu} p_{ik}^{(n)} + \sum_{n=0}^{N} {}_iP_{jk}^{(n)};$$

$$\sum_{n=0}^{N} p_{ji}^{(n)} = \sum_{\nu=1}^{N} f_{ji}^{(\nu)} \sum_{n=0}^{N-\nu} p_{ii}^{(n)};$$

where the first two follow trivially from $f_{i0}^{*}=1$, the last two from summing particular cases of (9.1). Combining these equations as indicated in (17), we can write its left member as follows:

$$\left\{\begin{aligned} &\sum_{\nu=1}^{N} f_{ji}^{(\nu)} \sum_{n=N-\nu+1}^{N} [p_{ii}^{(n)} - p_{ik}^{(n)}\, e_{ki}] + \sum_{\nu=N+1}^{\infty} f_{ji}^{(\nu)} \sum_{n=0}^{N} [p_{ii}^{(n)} - p_{ik}^{(n)}\, e_{ki}] + \\ &\qquad + \sum_{n=0}^{N} {}_iP_{jk}^{(n)}\, e_{ki}. \end{aligned}\right. \tag{18}$$

Now we have also, from summing particular cases of (9.2):

$$\sum_{n=0}^{N} p_{ii}^{(n)} = \sum_{\nu=0}^{N} p_{ik}^{(\nu)} \sum_{n=0}^{N-\nu} {}_k p_{ki}^{(n)} + \sum_{n=0}^{N} {}_k p_{ii}^{(n)} \leqq \sum_{\nu=0}^{N} p_{ik}^{(\nu)}\, e_{ki} + \sum_{n=0}^{N} {}_k p_{ii}^{(n)};$$

$$\sum_{n=0}^{N} p_{ik}^{(n)} = \sum_{\nu=0}^{N} p_{ii}^{(\nu)} \sum_{n=0}^{N-\nu} {}_i p_{ik}^{(n)} \leqq \sum_{\nu=0}^{N} p_{ii}^{(\nu)}\, e_{ik}.$$

Since $e_{ik}\, e_{ki}=1$ by Corollary 1 to Theorem 9.5, it follows from the inequalities above that

$$0 \leqq \sum_{n=0}^{N} p_{ii}^{(n)} - \sum_{n=0}^{N} p_{ik}^{(n)}\, e_{ki} \leqq \sum_{n=0}^{N} {}_k p_{ii}^{(n)} \leqq 1 + {}_k p_{ii}^{*} < \infty. \tag{19}$$

Consequently the second term in (18) converges to zero as $N \to \infty$. Furthermore,

$$[p_{ii}^{(n)} - p_{ik}^{(n)}\, e_{ki}] - [p_{ii}^{(n-1)} - p_{ik}^{(n-1)}\, e_{ki}] = [p_{ii}^{(n)} - p_{ii}^{(n-1)}] - [p_{ik}^{(n)} - p_{ik}^{(n-1)}]\, e_{ki}$$

converges to zero as $n \to \infty$, in either case of the hypothesis of the theorem. Hence the first term in (18) converges to zero as $N \to \infty$, by partial summation (ABEL's method). Letting $N \to \infty$ in (18) we therefore obtain (17). ▯

It follows at once from (19) and the divergence of $\sum_{n=0}^{\infty} p_{ii}^{(n)}$ that

$$\lim_{N\to\infty} \frac{\sum_{n=0}^{N} p_{ik}^{(n)}}{\sum_{n=0}^{N} p_{ii}^{(n)}} = \frac{1}{e_{ki}} = e_{ik} < \infty.$$

This and Theorem 5.3 yield a quick proof of (9.18) and hence also of Corollary 2 to Theorem 9.4 in a recurrent class.

Notes. Formulas (11) and (13) are due to HARRIS [1]; the rest is essentially in CHUNG [2] and [3] except for Theorem 8. Theorem 1 has been generalized here to allow for a taboo set. The proof of Theorem 3 is new; in CHUNG [2] certain other moments and their combinatorial relations are studied.

For empty H and $p=1$, Theorem 1 reduces to the result of KOLMOGOROV that all mean first entrance times are finite in a positive class. It is not exactly clear what property of the power n^p lies behind this theorem, although the method of proof permits trivial extensions.

The direct proof of Theorem 4 given in CHUNG [2] is incorrect because of non-absolute convergence in formula (15) there. Theorem 8 is taken from CHUNG [10], extending a result due to SPITZER on random walks; see NEVEU [4] for further generalization. The special case where all $e_{ki}=1$ is discussed in CHUNG [10].

§ 12. A random walk example

In this section we study in some detail a random walk scheme which is more general than the one we studied in § 5. First we develop a general method.

The taboo probabilities ${}_Hp_{ij}^{(n)}$ satisfy the obvious system of equations:

$$(1) \qquad {}_Hp_{ij}^{(n)} = \sum_{k \notin H}{}' p_{ik}\, {}_Hp_{kj}^{(n-1)}, \qquad n \geqq 2.$$

Summing over n we obtain

$$(2) \qquad {}_Hp_{ij}^{*} = \sum_{k \notin H} p_{ik}\, {}_Hp_{kj}^{*} + p_{ij}$$

provided that every ${}_Hp_{kj}^{*} < \infty$ and the series above converges. Next, multiplying (1) through by n, summing and using (2) we obtain

$$(3) \qquad {}_HM_{ij} = \sum_{k \notin H} p_{ik}\, {}_HM_{kj} + {}_Hp_{ij}^{*}$$

where

$$ {}_HM_{ij} = \sum_{n=1}^{\infty} n\, {}_Hp_{ij}^{(n)},$$

and both members of (3) may be ∞.

Specializing to the case where H consists of a single state j we see that $\{u_i, i \in \boldsymbol{I}\}$ where

$$u_i = \begin{cases} f_{ij}^{*} & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases}$$

is a solution of the system

$$u_i = \sum_l p_{il}\, u_l, \qquad i \neq j.$$

Specializing to the case where H consists of two states j and k where $j \neq k$ we see that $\{u_i, i \in \boldsymbol{I}\}$ where

$$u_i = \begin{cases} {}_kf_{ij}^{*} & \text{if } i \neq j \text{ or } k, \\ 1 & \text{if } i = j, \\ 0 & \text{if } i = k, \end{cases}$$

is a solution of the system

$$u_i = \sum_l p_{il}\, u_l, \qquad i \neq j \text{ or } k.$$

Furthermore $\{u_i, i \in \boldsymbol{I}\}$ where

$$u_i = \begin{cases} {}_km_{ij} & \text{if } i \neq j \text{ or } k, \\ 0 & \text{if } i = j \text{ or } k, \end{cases}$$

is a solution of the system

$$u_i = \sum_l p_{il}\, u_l + {}_k f^*_{ij}, \qquad i \neq j \text{ or } k.$$

Finally $\{u_i,\, i \in \boldsymbol{I}\}$ where

$$u_i = \begin{cases} {}_k m_{ij} + {}_j m_{ik} & \text{if } i \neq j \text{ or } k, \\ 0 & \text{if } i = j \text{ or } k, \end{cases}$$

is a solution of the system

$$u_i = \sum_l p_{il}\, u_l + {}_k f^*_{ij} + {}_j f^*_{ik}, \qquad i \neq j \text{ or } k.$$

If the state space is a recurrent class then the last system reduces to

$$u_i = \sum_l p_{il}\, u_l + 1.$$

In the first two cases the condition of finiteness and convergence is automatically satisfied while in the last two cases it must be ascertained. We are now going to apply these results to an example in which the above systems of equations with boundary values can be solved.

Consider a M.C. whose state space is the set of nonnegative integers. The initial distribution is arbitrary and the transition matrix is given as follows. Let $\alpha_0 = 1$, $0 < \alpha_j < 1$, $\beta_j = 1 - \alpha_j$;

$$p_{j,j+1} = \alpha_j, \qquad p_{j,j-1} = \beta_j, \qquad j \geqq 0.$$

Thus the generalization over the classical scheme (§ 5) lies in the fact that the transition probabilities at different states are generally different although from each state it is possible only to go to one of the two adjacent states at the next step. The condition $\alpha_0 = 1$ making 0 a reflecting barrier is not as restrictive as it may seem. By an obvious relabelling we could replace 0 by any negative integer and so achieve an apparent extension of the scheme. In particular, quantities like ${}_k f^*_{ij}$ and ${}_k m_{ij}$ which depend only on a finite portion of the transition matrix because of the special nature of the random walk will not be affected by any change outside this portion.

Let us first compute the f^*_{ij}. According to the general method they satisfy the following systems:

$$u_i = \alpha_i\, u_{i+1} + \beta_i\, u_{i-1}, \qquad 0 \leqq i < j;$$

$$u_j = 1.$$

Since $\alpha_0 = 1$ we have $u_0 = u_1$. Rewriting the system as

$$\alpha_i (u_{i+1} - u_i) = \beta_i (u_i - u_{i-1}),$$

we see that the unique solution is $u_i=\alpha$, $0\leqq i\leqq j$. Thus

$$f_{ij}^*=1, \qquad 0\leqq i<j. \tag{4}$$

This is to be expected since 0 is reflecting to the right.

To compute f_{ii}^* or f_{ji}^* we need ${}_kf_{ij}^*$. If $j<k$ they satisfy the following system:

$$u_i=\alpha_i u_{i+1}+\beta_i u_{i-1}, \qquad j<i<k;$$

$$u_j=1, \quad u_k=0.$$

This system can be solved easily by recursion and since the solution is unique it must be ${}_kf_{ij}^*$. We have, if $j<r<k$,

$$\alpha_r(u_{r+1}-u_r)=\beta_r(u_r-u_{r-1}),$$

$$u_{r+1}-u_r=\frac{\beta_r\cdots\beta_{j+1}}{\alpha_r\cdots\alpha_{j+1}}(u_{j+1}-1),$$

$$-1=\sum_{r=j}^{k-1}(u_{r+1}-u_r)=\sum_{r=j}^{k-1}\frac{\beta_r\cdots\beta_{j+1}}{\alpha_r\cdots\alpha_{j+1}}(u_{j+1}-1),$$

$$u_i-1=\sum_{r=j}^{i-1}(u_{r+1}-u_r)=\sum_{r=j}^{i-1}\frac{\beta_r\cdots\beta_{j+1}}{\alpha_r\cdots\alpha_{j+1}}(u_{j+1}-1),$$

where we note, conformably to established convention regarding an empty product,

$$\frac{\beta_r\cdots\beta_{j+1}}{\alpha_r\cdots\alpha_{j+1}}=1$$

if $r=j$ in the above. Setting

$$\varrho_0=1, \qquad \varrho_r=\frac{\beta_1\cdots\beta_r}{\alpha_1\cdots\alpha_r}$$

we obtain

$${}_kf_{ij}^*=\frac{\sum\limits_{r=i}^{k-1}\varrho_r}{\sum\limits_{r=j}^{k-1}\varrho_r}, \qquad j<i<k.$$

It is worthy of mention that the expression on the right side depends only on α_r, β_r with $j<r<k$, as alluded to earlier. Similarly or by a reversal argument we have

$${}_jf_{ik}^*=\frac{\sum\limits_{r=j}^{i-1}\varrho_r}{\sum\limits_{r=j}^{k-1}\varrho_r}, \qquad j<i<k. \tag{5}$$

The cases $i=j$ and $i=k$ are trivial. We have, if $j<k$,

$$ {}_j f^*_{jk} = \alpha_j \, {}_j f^*_{j+1,k} = \frac{\alpha_j \varrho_j}{\sum\limits_{r=j}^{k-1} \varrho_r}, $$

$$ {}_k f^*_{kj} = \beta_k \, {}_k f^*_{k-1,j} = \frac{\varrho_{k-1} \beta_k}{\sum\limits_{r=j}^{k-1} \varrho_r}, $$

$$ {}_k f^*_{jj} = \alpha_j \, {}_k f^*_{j+1,j} + \beta_j \, {}_k f^*_{j-1,j} = \frac{\alpha_j \sum\limits_{r=j+1}^{k-1} \varrho_r}{\sum\limits_{r=j}^{k-1} \varrho_r} + \beta_j = 1 - {}_j f^*_{jk}. $$

Of course the fact ${}_k f^*_{jj} + {}_j f^*_{jk} = 1$ in this scheme whether the class is recurrent or not can be foreseen.

We can now decide when the random walk is recurrent. By Theorem 11.5 with $p=0$ we have

$$ f^*_{10} = \lim_{k\to\infty} {}_k f^*_{10} = \lim_{k\to\infty} \frac{\sum\limits_{r=1}^{k-1} \varrho_r}{\sum\limits_{r=0}^{k-1} \varrho_r}. $$

Hence $f^*_{10}=1$ if and only if

$$ \sum_r \varrho_r = \sum_r \frac{\beta_1 \cdots \beta_r}{\alpha_1 \cdots \alpha_r} = \infty. \tag{6} $$

Under the condition (6) we have by (4), $f^*_{00} \geqq f^*_{01} f^*_{10} = 1$. Therefore by Theorem 5.6, (6) is a necessary and sufficient condition for the random walk to be recurrent.

In case of nonrecurrence, we have by Theorem 11.5,

$$ f^*_{jj} = \lim_{k\to\infty} {}_k f^*_{jj} = 1 - \frac{\alpha_j \varrho_j}{\sum\limits_{r=j}^{\infty} \varrho_r}. $$

Moreover, if $i>j$,

$$ f^*_{ij} = \lim_{k\to\infty} {}_k f^*_{ij} = \frac{\sum\limits_{r=i}^{\infty} \varrho_r}{\sum\limits_{r=j}^{\infty} \varrho_r}. $$

In case of recurrence, we have by Theorem 9.5,

$$ \lim_{N\to\infty} \frac{\sum\limits_{n=1}^{N} p^{(n)}_{kk}}{\sum\limits_{n=1}^{N} p^{(n)}_{00}} = {}_0 p^*_{0k} = \frac{{}_0 f^*_{0k}}{{}_k f^*_{k0}} = \frac{1}{\varrho_{k-1} \beta_k} = \frac{1}{\alpha_k \varrho_k}, \qquad k \geqq 1. $$

Hence by the Corollary to Theorem 9.6, the recurrent class of the M.C. is positive if and only if

$$(7) \qquad \sum_k \frac{1}{\alpha_k \varrho_k} = \sum_k \frac{\alpha_1 \cdots \alpha_{k-1}}{\beta_1 \cdots \beta_{k-1} \beta_k} < \infty .$$

Under the condition (7) the preceding relation reduces to

$$\frac{m_{00}}{m_{kk}} = \frac{1}{\alpha_k \varrho_k}$$

and consequently by Theorem 7.1,

$$m_{00} = 1 + \sum_{k=1}^{\infty} \frac{1}{\alpha_k \varrho_k} .$$

Thus all m_{kk} are evaluated.

To compute the other m_{ij} in the positive-recurrent case we first compute ${}_k m_{ij} + {}_j m_{ik}$ for $j < i < k$. According to the general method they satisfy the system

$$u_i = \alpha_i u_{i+1} + \beta_i u_{i-1} + 1 , \qquad j < i < k .$$

This system can be solved in the same way as detailed above. We obtain

$$(8) \qquad {}_n m_{ij} + {}_j m_{ik} = \sum_{s=i}^{k-1} \varrho_s \sum_{r=j+1}^{s} \frac{1}{\alpha_r \varrho_r} - \frac{\sum\limits_{s=i}^{k-1} \varrho_s}{\sum\limits_{s=j}^{k-1} \varrho_s} \sum_{s=j}^{k-1} \varrho_s \sum_{r=j+1}^{s} \frac{1}{\alpha_r \varrho_r} , \qquad j < i < k .$$

By Theorem 11.5 and its Corollary, we have since $f_{kj}^* = 1$,

$$\begin{aligned} m_{ij} &= \lim_{k \to \infty} ({}_k m_{ij} + {}_j m_{ik}) \\ &= \lim_{k \to \infty} \frac{\sum\limits_{s=i}^{k-1} \varrho_s \sum\limits_{r=j+1}^{s} \frac{1}{\alpha_r \varrho_r}}{\sum\limits_{s=j}^{k-1} \varrho_s} \sum_{s=j}^{i-1} \varrho_s - \lim_{k \to \infty} \frac{\sum\limits_{s=i}^{k-1} \varrho_s}{\sum\limits_{s=j}^{k-1} \varrho_s} \sum_{s=j}^{i-1} \varrho_s \sum_{r=j+1}^{s} \frac{1}{\alpha_r \varrho_r} \end{aligned}$$

after some simplification. Now since $\sum\limits_s \varrho_s = \infty$, $\lim\limits_{s \to \infty} A_s = A$ implies that

$$\lim_{k \to \infty} \frac{\sum\limits_{s=i}^{k-1} A_s \varrho_s}{\sum\limits_{s=i}^{k-1} \varrho_s} = A .$$

Applying this result we obtain

$$m_{ij} = \sum_{r=j+1}^{\infty} \frac{1}{\alpha_r \varrho_r} \sum_{s=j}^{i-1} \varrho_s - \sum_{s=j}^{i-1} \varrho_s \sum_{r=j+1}^{s} \frac{1}{\alpha_r \varrho_r}$$
$$= \sum_{s=j}^{i-1} \varrho_s \sum_{r=s+1}^{\infty} \frac{1}{\alpha_r \varrho_r}, \qquad j < i.$$

We see that $m_{ij} = \sum_{k=j}^{i-1} m_{k+1,k}$ which is obvious directly.

We can now compute m_{ji} from (11.11):

$$m_{jj} = {}_j f^*_{ji} (m_{ji} + m_{ij}) = \frac{\alpha_j \varrho_j}{\sum_{s=j}^{i-1} \varrho_s} \left(m_{ji} + \sum_{s=j}^{i-1} \varrho_s \sum_{r=s+1}^{\infty} \frac{1}{\alpha_r \varrho_r} \right)$$

whence we obtain after some simplification,

$$m_{ji} = \sum_{s=j}^{i-1} \varrho_s \left(1 + \sum_{r=1}^{s} \frac{1}{\alpha_r \varrho_r} \right), \qquad j < i.$$

We can go further and compute ${}_k m_{ij}$. To do this we need only solve the system

$$u_i = \alpha_i u_{i+1} + \beta_i u_{i-1} + {}_k f^*_{ij}, \qquad j < i < k.$$

The result is analogous to (8):

$$(9) \qquad {}_k m_{ij} = \sum_{s=i}^{k-1} \varrho_s \sum_{r=j+1}^{s} \frac{{}_k f^*_{rj}}{\alpha_r \varrho_r} - \frac{\sum_{s=i}^{k-1} \varrho_s}{\sum_{s=j}^{k-1} \varrho_s} \sum_{s=j}^{k-1} \varrho_s \sum_{r=j+1}^{s} \frac{{}_k f^*_{rj}}{\alpha_r \varrho_r}, \qquad j < i < k.$$

Subtracting (9) from (8) we see that ${}_j m_{ik}$ is obtained from the right member of (9) after replacing ${}_k f^*_{rj}$ by ${}_j f^*_{rk}$.

To illustrate the above results in a particular case, we recall Example 4 of § 5. The f^*_{il} there is seen to be ${}_0 f^*_{il}$ here, $0 < i < l$, given in (5) with $\varrho_r = \left(\frac{1-p}{p} \right)^r$. This yields the value indicated there.

Notes. The random walk scheme is treated in HARRIS [1] by an unnecessarily devious method. The method employed here, that of difference equations, is of course classical (see e.g. USPENSKY [1]; Chapter 8). Our results go beyond those given by HARRIS. For results under special forms of the α_n see HARRIS's paper and also HODGES and ROSENBLATT [1].

§ 13. System theorems

In the following we shall frequently put

$$x(t, \omega) = x_t(\omega).$$

Let α be a nonnegative integer-valued random variable with domain of finiteness $\Delta\in\mathscr{F}$, $\boldsymbol{P}\{\Delta\}>0$. For each $t\geqq 0$,

$$(1)\qquad y_t(\omega)=y(t,\omega)=x\big(\alpha(\omega)+t,\omega\big)$$

is the value at ω of a random variable y_t whose domain of definition and finiteness is Δ; indeed we have

$$\{\omega\colon y_t(\omega)=i\}=\bigcup_{n=0}^{\infty}\{\omega\colon \alpha(\omega)=n;\ x_{n+t}(\omega)=i\}.$$

In particular we have for any $\Lambda\in\mathscr{F}$:

$$(2)\qquad \boldsymbol{P}\{\Lambda;\, y_0(\omega)=i\}=\sum_{n=0}^{\infty}\boldsymbol{P}\{\Lambda;\, \alpha(\omega)=n;\ x_n(\omega)=i\},\qquad i\in\boldsymbol{I}.$$

We define $\Delta\mathscr{F}$ to be the (augmented) Borel field of those sets of $\mathscr{F}$ contained in Δ; for a set Λ in $\Delta\mathscr{F}$ the appropriate probability measure is the conditional probability

$$\boldsymbol{P}_\Delta(\Lambda)=\boldsymbol{P}(\Lambda\,|\,\Delta)=\frac{\boldsymbol{P}(\Lambda)}{\boldsymbol{P}(\Delta)}.$$

The triple $(\Delta,\Delta\mathscr{F},\boldsymbol{P}(\cdot\,|\,\Delta))$ is called *the probability triple obtained by reducing* $(\Omega,\mathscr{F},\boldsymbol{P})$ *to* Δ. Clearly the $\{y_t,t\geqq 0\}$ defined in (1) is a stochastic process on the reduced triple. A significant case arises when α satisfies the following condition: for every $n\geqq 0$,

$$\{\omega\colon \alpha(\omega)=n\}\in\mathscr{F}\{x_s,\, s\leqq n\}.$$

A nonnegative, integer-valued random variable satisfying this condition will be called *optional* (relative to the M.C.). If α is optional then the *pre-α field* $\mathscr{F}\{x_s,\, s\leqq\alpha\}$ is defined to be the Borel field of all sets Λ in $\Delta\mathscr{F}$ such that

$$\Lambda\cap\{\omega\colon \alpha(\omega)\leqq n\}\in\mathscr{F}\{x_s,\, s\leqq n\},\qquad n\geqq 0.$$

The intuitive background of an optional time may be explained as follows. Instead of observing the entire process x_t it is frequently of interest to consider its evolution only before, at or after a certain time α to be chosen according to fortuitous circumstances. In general there is no reason why such a choice should not depend on the evolution of the process after the chosen time, for example we may define α to be the time immediately preceding (namely, one time-unit before) the first entrance into a given state. The optional condition however rules this out by requiring that the choice must be determined only by events which happened no later than the chosen time, as if we were actually watching the gradual unravelling of the system and must decide on the choice of α without the advantage of foreseeing the future. The situation is descriptive of the gambler awaiting his turn to bet on a game in progress according to a winning system he has devised. Hence theorems of the type given below are called *system theorems.*

If α is optional the stochastic process $\{y_t, t \geqq 0\}$ defined in (1) is called the *post-α process* (relative to the M.C.), and the Borel field $\mathscr{F}\{y_t, t \geqq 0\}$, to be denoted also by $\mathscr{F}\{x_t, t \geqq \alpha\}$, is called the post-$\alpha$ field.

Theorem 1. Let α be an optional random variable with the domain of finiteness Δ where $\boldsymbol{P}(\Delta) > 0$. The post-$\alpha$ process $\{y_t, t \geqq 0\}$ is then a M.C. on $(\Delta, \Delta\mathscr{F}, \boldsymbol{P}(\cdot \mid \Delta))$ with state space $\boldsymbol{J} \subset \boldsymbol{I}$, the initial distribution given in (2) with $\Lambda = \Delta$, and the transition matrix (p_{ij}) restricted to $\boldsymbol{J}$. Moreover if $\Lambda \in \mathscr{F}\{x_s, s \leqq \alpha\}$ and $0 = t_0 < t_1 < \cdots < t_l$; $i_0, i_1, \ldots, i_l \in \boldsymbol{I}$, then

$$(3) \qquad \boldsymbol{P}\{\Lambda; y(t_\nu, \omega) = i_\nu, \ 0 \leqq \nu \leqq l\} = \boldsymbol{P}\{\Lambda; y(0, \omega) = i_0\} \prod_{\nu=0}^{l-1} p_{i_\nu i_{\nu+1}}^{(t_{\nu+1} - t_\nu)}.$$

Proof. By the definition of $\mathscr{F}\{x_s, s \leqq \alpha\}$ we have for every $n \geqq 0$,

$$\Lambda \cap \{\omega : \alpha(\omega) = n\} = \Lambda_n \cap \{\omega : \alpha(\omega) = n\}$$

where $\Lambda_n \in \mathscr{F}\{x_s, s \leqq n\}$. Let $\mathsf{M} \in \mathscr{F}\{y_s, s \geqq 0\}$, then similarly

$$\mathsf{M} \cap \{\omega : \alpha(\omega) = n\} = \mathsf{M}_n \cap \{\omega : \alpha(\omega) = n\}$$

where $\mathsf{M}_n \in \mathscr{F}\{x_s, s \geqq n\}$. Hence we have

$$(4) \quad \begin{cases} \boldsymbol{P}(\Lambda \mathsf{M}) = \sum\limits_{n=0}^{\infty} \sum\limits_{i \in \boldsymbol{I}} \boldsymbol{P}\{\Lambda_n; \alpha(\omega) = n; x_n(\omega) = i; \mathsf{M}_n\} \\ \qquad = \sum\limits_{n=0}^{\infty} \sum\limits_{i \in \boldsymbol{I}} \boldsymbol{P}\{\Lambda; \alpha(\omega) = n; x_n(\omega) = i\} \boldsymbol{P}\{\mathsf{M}_n \mid x_n(\omega) = i\}. \end{cases}$$

Here we have used the optionality of α in order to apply the Markov property. In particular if

$$\mathsf{M} = \bigcap_{\nu=1}^{l} \{\omega : y(t_\nu, \omega) = i_\nu\}$$

then

$$\mathsf{M}_n = \bigcap_{\nu=1}^{l} \{\omega : x(n + t_\nu, \omega) = i_\nu\}$$

and

$$\boldsymbol{P}\{\mathsf{M}_n \mid x_n(\omega) = i_0\} = \prod_{\nu=0}^{l-1} p_{i_\nu i_{\nu+1}}^{(t_{\nu+1} - t_\nu)} = P_{i_0},$$

say, which does not depend on n. Substituting into (4) and using (2) we obtain

$$\boldsymbol{P}\{\Lambda; y(t_\nu, \omega) = i_\nu, 1 \leqq \nu \leqq l\} = \sum_{i_0 \in \boldsymbol{I}} \boldsymbol{P}\{\Lambda; y(0, \omega) = i_0\} P_{i_0}.$$

Since $\{\omega : y(0, \omega) = i_0\} \in \mathscr{F}\{x_s, s \leqq \alpha\}$, we may replace the set Λ in the equation above by $\Lambda \cap \{\omega : y(0, \omega) = i_0\}$. The resulting equation reduces to (3). In particular if $\Lambda = \Delta$ we see that $\{y_t, t \geqq 0\}$ is a M.C. with the stated properties. ▯

The following special case of Theorem 1 will be used.

Theorem 2. If in addition we have

$$(5) \qquad \boldsymbol{P}_{\Delta}\{y(0, \omega) = i_0\} = 1$$

for a certain i_0, then

$$\boldsymbol{P}(\Lambda \mathsf{M}) = \boldsymbol{P}(\Lambda)\, \boldsymbol{P}_{\Delta}(\mathsf{M});$$

in particular,

$$\boldsymbol{P}\{\Lambda;\, y(t_\nu, \omega) = i_\nu,\ 1 \leqq \nu \leqq l\} = \boldsymbol{P}(\Lambda)\, P_{i_0}.$$

This is often expressed by saying that "the process starts from scratch" from the state i_0 at time α. The part of the Theorem 2 which asserts independence will now be generalized. The intersection $\mathscr{F}_1 \cap \mathscr{F}_2$ of two fields $\mathscr{F}_1$ and $\mathscr{F}_2$ is of course the field of sets belonging to both.

Theorem 3. Let each α_n, $1 \leqq n \leqq l$, be an optional random variable with domain of finiteness Δ_n such that $\Delta_n \supset \Delta_{n+1}$ and $\alpha_n < \alpha_{n+1}$ on Δ_{n+1}. Suppose further that for each n there is a state i_n such that

$$\boldsymbol{P}_{\Delta_n}\{x(\alpha_n(\omega), \omega) = i_n\} = 1.$$

Let Λ_n, $1 \leqq n \leqq l+1$, be ω-sets such that

$$\Lambda_n \in \mathscr{F}\{x_s,\, s \geqq \alpha_{n-1}\} \cap \mathscr{F}\{x_s,\, s \leqq \alpha_n\}$$

where $\alpha_0 = 0$, $\alpha_{l+1} = \infty$. Then

$$\boldsymbol{P}\left\{\bigcap_{n=1}^{l+1} \Lambda_n\right\} = \frac{\prod\limits_{n=1}^{l+1} \boldsymbol{P}(\Lambda_n)}{\prod\limits_{n=1}^{l} \boldsymbol{P}(\Delta_n)}.$$

Proof. By a repeated application of Theorem 2, we have, writing $\Delta_0 = \Omega$,

$$\boldsymbol{P}\left\{\bigcap_{n=1}^{l+1} \Lambda_n\right\} = \prod_{n=1}^{l+1} \boldsymbol{P}_{\Delta_0 \Delta_1 \ldots \Delta_{n-1}}(\Lambda_n).$$

But $\Delta_0 \Delta_1 \ldots \Delta_{n-1} = \Delta_{n-1}$, whence the theorem. ▯

Corollary. If in addition each α_n is finite with probability one, then the sets Λ_n are independent.

In the definition of the $\{y_t\}$ process the random variable α serves as an optional starting time. We shall now consider an optional stopping time. Let $\alpha'(\omega)$ be the smallest value of $n \geqq 1$ such that $x_n(\omega) = i$ if such a value exists; otherwise let $\alpha'(\omega) = \infty$. Then α' is a random variable whose domain of definition is Ω. We define

$$z(t, \omega) = \begin{cases} x(t, \omega) & \text{if } 1 \leqq t < \alpha'(\omega); \\ x(\alpha'(\omega), \omega) = i & \text{if } t \geqq a'(\omega). \end{cases}$$

The following theorem is an immediate consequence of the definition.

Theorem 4. The process $\{z_t, t \geqq 1\}$ defined above is a M.C. on $(\Omega, \mathscr{F}, \boldsymbol{P})$ of which the state space $\boldsymbol{K}$ is a subset of $\boldsymbol{I}$; the initial distribution $\{q_k, k \in \boldsymbol{K}\}$ is that of x_1, namely

$$q_k = \sum_{j \in \boldsymbol{I}} p_j \, p_{jk};$$

and the transition matrix (q_{jk}), $j, k \in \boldsymbol{K}$, is given by

$$q_{jk} = \begin{cases} p_{jk} & \text{if } j \neq i \\ \delta_{jk} & \text{if } j = i. \end{cases}$$

Corollary. The n-step transition matrix is given by

$$\begin{aligned} q_{jk}^{(n)} &= {}_i p_{jk}^{(n)} && \text{if } j \neq i,\ k \neq i; \\ q_{ji}^{(n)} &= \sum_{\nu=1}^{n} f_{ji}^{(\nu)} && \text{if } j \neq i; \\ q_{ii}^{(n)} &= 1. \end{aligned}$$

The following combination of Theorems 2 and 4 will be used shortly.

Theorem 5. Let α be an optional random variable with domain of finiteness Δ and satisfying condition (5). Let α' be defined on Δ as follows: $\alpha'(\omega)$ is the least value of $n > \alpha(\omega)$ such that $x_n(\omega) = i$ if such a value exists; otherwise $\alpha'(\omega) = \infty$. Define a process $\{\xi_t, t \geqq 0\}$ on Δ as follows:

$$\xi(t, \omega) = \begin{cases} x(\alpha(\omega) + t, \omega) & \text{if } 1 \leqq t < \alpha'(\omega) - \alpha(\omega); \\ x(\alpha'(\omega)) = i & \text{if } t \geqq \alpha'(\omega) - \alpha(\omega). \end{cases} \tag{6}$$

Let $\Lambda \in \mathscr{F}\{x_s, s \leqq \alpha\}$; let $0 = t_0 < t_1 < \cdots < t_m$; $i_0 = i$ and $i_1, \ldots, i_m$ be a sequence of states such that for some l, $1 \leqq l \leqq m-1$, we have $i_\nu \neq i$ if $1 \leqq \nu < l$ and $i_\nu = i$ if $l \leqq \nu \leqq m$. Then we have

$$\boldsymbol{P}\{\Lambda;\ \xi(t_\nu, \omega) = i_\nu,\ 0 \leqq \nu \leqq m\} = \boldsymbol{P}(\Lambda) \prod_{\nu=0}^{l} {}_i p_{i_\nu i_{\nu+1}}^{(t_{\nu+1} - t_\nu)}.$$

Moreover if the sequence $i_1, \ldots, i_m$ does not have the prescribed property then the above probability is 0.

The fact that the $\{\xi_t\}$ process is a Markov chain is less interesting than its transition properties described above. In fact, we are interested in this process only for the values of t satisfying $1 \leqq t \leqq \alpha'(\omega) - \alpha(\omega)$. The choice of the value i for $t > \alpha'(\omega) - \alpha(\omega)$, making it thereby Markovian, is merely convenient.

Notes. The concept of a system theorem is due to Doob (see [4, pp. 144, 300]). Of course in the case discussed in this section the results are very simple and intuition is easily rigorized. The point is that the defining Markov property is stated only for constant parameter values, for which random variables cannot be substituted without proper justification. Such justification is no longer trivial in Part II, where corresponding results for a continuous parameter M.C. will be given.

§ 14. Functionals and associated random variables

In this and the next two sections we consider a M.C. $\{x_n, n \geqq 0\}$ on $(\Omega, \mathscr{F}, \boldsymbol{P})$ with the state space $\boldsymbol{I}$ forming a recurrent class. Thus for almost all $\omega \in \Omega$, the sequence $\{x_n(\omega), n \geqq 0\}$ contains an infinite number of every state $i \in \boldsymbol{I}$. By discarding a set of probability zero we may and we shall assume that this is indeed true for all $\omega \in \Omega$ without exception.

For each $i \in \boldsymbol{I}$ and each $\omega \in \Omega$ let then

$$\tau_1(i;\omega) < \cdots < \tau_\nu(i;\omega) < \cdots$$

be the increasing infinite sequence of all values of $n \geqq 1$ for which $x_n(\omega) = i$. It is easy to see that the ω-function

$$\tau_\nu(i) = \tau_\nu(i;\cdot)$$

is an optional random variable with Ω as domain of definition and finiteness. This random variable is called the *ν-th entrance time of the state i*. The half-open random interval $I_\nu(i;\omega) = [\tau_\nu(\omega), \tau_{\nu+1}(\omega))$ is called the *ν-th return interval of the state i*; and the random variable

$$\varrho_\nu(i) = \tau_{\nu+1}(i) - \tau_\nu(i)$$

the ν-th return time. It is clear from the Corollary to Theorem 13.3 that they have the common distribution given by

$$\boldsymbol{P}\{\varrho_\nu(\omega) = n\} = f_{ii}^{(n)}, \quad n \geqq 1.$$

It is convenient to set $\tau_0(i;\omega) = 0$. For each $n \geqq 0$ and $\omega \in \Omega$ there is a unique nonnegative integer $l(i; n, \omega)$ satisfying

$$(1) \qquad \tau_{l(i;n,\omega)}(i;\omega) \leqq n < \tau_{l(i;n,\omega)+1}(i;\omega).$$

The ω-functions $l(i; n, \cdot)$, $\tau_{l(i;n,\cdot)}(i;\cdot)$ and $\tau_{l(i;n,\cdot)+1}(i;\cdot)$ are random variables. For the sake of brevity we denote, for a fixed i, the two end terms in (1) by $\tau(n,\omega)$ and $\tau'(n,\omega)$ respectively.

We state our first theorem only for a class of period 1 since we do not need it in its precise form.

Theorem 1. If the period of the class $\boldsymbol{I}$ is 1, then for each $i \in \boldsymbol{I}$,

$$\lim_{n\to\infty} \boldsymbol{P}\{n-\tau(n,\omega)=t\}=\pi_i \sum_{s=t+1}^{\infty} f_{ii}^{(s)}, \qquad t \geqq 0;$$

$$\lim_{n\to\infty} \boldsymbol{P}\{\tau'(n,\omega)-n=t\}=\pi_i \sum_{s=t}^{\infty} f_{ii}^{(s)}, \qquad t \geqq 1;$$

$$\lim_{n\to\infty} \boldsymbol{P}\{\tau'(n,\omega)-\tau(n,\omega)=t\}=\pi_i\, t\, f_{ii}^{(t)}, \quad t \geqq 1;$$

where π_i is m_{ii}^{-1} or 0 according as $\boldsymbol{I}$ is positive or null (see end of §6).

Proof. It is easy to see that

$$(2) \quad \begin{cases} \boldsymbol{P}\{n-\tau(n,\omega)=t\}=p_i^{(n-t)} \sum\limits_{s=t+1}^{\infty} f_{ii}^{(s)}, & \text{if } t<n; \\ \boldsymbol{P}\{\tau'(n,\omega)-n=t\}=p_i^{(n+t)} - \sum\limits_{s=1}^{t-1} p_i^{(n+s)} f_{ii}^{(t-s)}, & \text{if } 1 \leqq n; \\ \boldsymbol{P}\{\tau'(n,\omega)-\tau(n,\omega)=t\}= \sum\limits_{s=n-t+1}^{n} p_i^{(s)} f_{ii}^{(t)}, & \text{if } t \leqq n. \end{cases}$$

Since $p_i^{(n)}=\sum_j p_j\, p_{ji}^{(n)} \to \pi_i$ by Theorem 6.1 the results follow immediately as $n\to\infty$. ▯

It is noteworthy that the distribution of $\tau'(n,\cdot)-\tau(n,\cdot)$ is nott hat of ϱ_ν.

Theorem 2. If $\boldsymbol{I}$ is a positive class then the three probabilities

$$\boldsymbol{P}\{n-\tau(n,\omega)\geqq t\}, \qquad \boldsymbol{P}\{\tau'(n,\omega)-n\geqq t\}, \qquad \boldsymbol{P}\{\tau'(n,\omega)-\tau(n,\omega)\geqq t\}$$

all tend to 0 as $t\to\infty$, uniformly with respect to n.

Proof. Whatever the period d of $\boldsymbol{I}$, we have from (2),

$$\overline{\lim_{n\to\infty}}\, \boldsymbol{P}\{n-\tau(n,\omega)\geqq t\} \leqq d\pi_i \sum_{s=t}^{\infty} \sum_{\nu=t+1}^{\infty} f_{ii}^{(\nu)} = d\pi_i \sum_{\nu=i+1}^{\infty} (\nu-t)\, f_{ii}^{(\nu)}.$$

If $\boldsymbol{I}$ is positive then $\sum_{\nu=1}^{\infty} \nu\, f_{ii}^{(\nu)}=\pi_i^{-1}<\infty$. Hence given any $\varepsilon>0$ there exist $t_0(\varepsilon)$ and $n_0(\varepsilon)$ such that $t>t_0$ and $n>n_0$ imply

$$\boldsymbol{P}\{n-\tau(n,\omega)\geqq t\}<\varepsilon.$$

Thus the above inequality holds for all if $t>\max(t_0, n_0)$. This proves the uniform convergence of the first probability in the theorem. The rest is similar. ▯

We now introduce the concept of a functional of a M.C. Let f be a function from $\boldsymbol{I}$ to the open real line $(-\infty, +\infty)$. The stochastic process $\{y_n, n \geqq 0\}$ on $(\Omega, \mathscr{F}, \boldsymbol{P})$ defined by

$$y_n = f(x_n), \qquad n \geqq 0$$

is called *a functional of the* $\{x_n\}$ *process*. Clearly each y_n is a discrete random variable. The state space of the $\{y_n\}$ process is a denumerable set of real numbers which is denoted by $f(\boldsymbol{I})$. It is easy to see that a functional of a M.C. is in general not a M.C. However, the probability-average properties of the sequences $\{x_n(\omega), n \geqq 0\}$ extend easily to those of the "weighted" sequences $\{y_n(\omega), n \geqq 0\}$ under broad conditions. It is the object of these sections to study these properties, or more precisely, the asymptotic properties of the partial sums

$$S_n(\omega) = \sum_{s=0}^{n} y_s(\omega)$$

as $n \to \infty$.

Let us first introduce some notation to be used henceforth. Let i be a fixed state; *we shall frequently omit i from the notation* where there is no danger of ambiguity. Define

$$Y_\nu(i; \omega) = \sum_{n=\tau_\nu(i;\omega)}^{\tau_{\nu+1}(i;\omega)-1} y_n(\omega) = \sum_{n \in I_\nu(i;\omega)} y_n(\omega).$$

The ω-function $Y_\nu(i) = Y_\nu(i; \cdot)$ is a random variable that is discrete and finite with probability one. In view of (1) we have

$$S_n(\omega) = \sum_{s=0}^{\tau_1(i;\omega)-1} y_s(\omega) + \sum_{\nu=1}^{l(i;n,\omega)-1} Y_\nu(i;\omega) + \sum_{s=\tau_{l(i;n,\omega)}(i;\omega)}^{n} y_s(\omega),$$

provided that $n \geqq \tau_1(i; \omega)$. In general we put, omitting i and ω from the notation:

$$Y'(n) = \sum_{s=0}^{\min(n, \tau_1 - 1)} y_s, \qquad Y''(n) = \begin{cases} 0 & \text{if } l(n) = 0, \\ \sum_{s=\tau_l}^{n} y_s & \text{if } l(n) \geqq 1; \end{cases}$$

and observing the usual convention regarding empty summation, we may write

$$S_n = Y'(n) + \sum_{\nu=1}^{l(n)-1} Y_\nu + Y''(n). \tag{3}$$

This basic formula will be referred to in the sequel as the *dissection formula*.

The first term $Y'(n)$ is easy to handle. It is seen that

$$|Y'(n)| \leqq \sum_{s=0}^{\tau_1 - 1} |y_s| \tag{4}$$

where the right side does not depend on n. Thus $Y'(n)$ is dominated by a fixed random variable that is finite with probability one.

As for the sum (3), we state the following leading theorem.

Theorem 3. The random variables $\{Y_\nu(i), \nu \geqq 1\}$ are independent and identically distributed. Furthermore let α be a positive integer-valued optional random variable and $\tau_{\alpha+\nu} = \tau_{\alpha(\omega)+\nu}(i; \omega)$ for integer $\nu \geqq 0$. Then the random variables $\{Y_{\alpha+\nu}(i); \nu \geqq 0\}$ are independent and identically distributed with the same distribution as each Y_ν.

Proof. We first apply Theorem 13.3 with $\alpha_n = \tau_n(i)$, $\Delta_n = \Omega$; the conditions there are obviously satisfied since a recurrent class is in question. For every real number c we have

$$\{\omega: Y_\nu(i; \omega) = c\} \in \mathscr{F}\{x_s, s \geqq \tau_\nu(i)\} \cap \mathscr{F}\{x_s, s \leqq \tau_{\nu+1}(i)\}.$$

Hence the random variables $\{Y_\nu(i), \nu \geqq 1\}$ are independent by the Corollary to Theorem 13.3. Writing

$$\hat{f}(j) = (1 - \delta_{ij}) f(j)$$

we see that in terms of the $\{\xi_n\}$ process defined in (13.6) we have

$$Y_\nu(i; \omega) = f(i) + \sum_{n=1}^{\infty} \hat{f}[\xi_n(\omega)], \tag{5}$$

where the series terminates with probability one owing to the recurrence of i. According to Theorem 13.5 the distribution properties of the $\{\xi_n\}$ process are the same regardless of ν. The same is therefore true of $Y_\nu(i)$.

It is easy to verify that $\tau_{\alpha+\nu}$ is optional for each ν. Hence the second assertion in Theorem 3 is an immediate generalization of the first and is proved in the same way. ▯

Our next order of business concerns the moments of $Y_\nu(i)$. We shall write $Y(i)$ for any one of the $Y_\nu(i)$, $\nu \geqq 1$, where ν is immaterial; thus we define

$$\mu_i^{(p)} = \boldsymbol{E}\{Y(i)^p\}, \qquad p \geqq 0;$$

and write $\mu_i = \mu_i^{(1)}$. As usual we say the moment of order p exists in case the absolute moment of order p is finite. The following theorem overlaps Theorem 11.1 and shows that the solidarity among the states of a recurrent class carries over even to the functional.

Theorem 4. If $\mu_i^{(p)}$ exists for any i, then it exists for every i.

Proof. Consider the sequence of random variables $\{\tau_n(i), n \geqq 1\}$ and also the similar sequence $\{\tau_n(j), n \geqq 1\}$ where $i \neq j$. Let $\nu(\omega)$ be the smallest value of n such that $\tau_n(j; \omega) > \tau_1(i; \omega)$; $N(\omega)$ the smallest value of n such that $\tau_n(i;\omega)>\tau_{\nu(\omega)+1}(j;\omega)$. Thus we have, in the notation of ω-functions,

$$\tau_1(i)<\tau_\nu(j)<\tau_{\nu+1}(j)<\tau_N(i).$$

Clearly ν, N and the last three τ's above are random variables. Furthermore, let $N'(\omega)$ be the smallest value of $s\geqq 1$ such that $x_n(\omega)=j$ for at least one n in the interval $I_s(i;\omega)$; $N''(\omega)$ the smallest value of $s\geqq 1$ such that $x_n(\omega)=j$ for at least one n in the interval $I_{N'(\omega)+s}(i;\omega)$. Observe that $\tau_{N'+1}$ and $\tau_{N'+N''+1}$ are optional random variables. It is obvious that

$$N\leqq N'+N''+1. \tag{6}$$

The random variables N' and N'' are introduced to facilitate computations. Let $\delta_j(\cdot)$ be the Kronecker delta function for a fixed j, namely $\delta_j(i)=\delta_{ij}$. Setting

$$A_s(\omega)=\sum_{n\in I_s(i;\,\omega)}\delta_j[x_n(\omega)]$$

we have

$$\boldsymbol{P}\{N'(\omega)=t\}=\boldsymbol{P}\{A_s(\omega)=0,\ 1\leqq s<t;\ A_t(\omega)>0\}.$$

Applying Theorem 3 we obtain

$$\boldsymbol{P}\{N'(\omega)=t\}=\boldsymbol{P}\{A_1(\omega)=0\}^{t-1}\boldsymbol{P}\{A_1(\omega)>0\}.$$

The event $A_1(\omega)=0$ is the event that $x_n(\omega)\neq j$ between the first and second entrances at i; hence

$$\boldsymbol{P}\{A_1(\omega)=0\}={}_jf_{ii}^*,\qquad \boldsymbol{P}\{A_1(\omega)>0\}=1-{}_jf_{ii}^*.$$

Thus the random variable N' has a geometric distribution and we have for every $p\geqq 0$,

$$\boldsymbol{E}\{(N')^p\}<\infty. \tag{7}$$

Now consider the sum

$$W=\sum_{n=1}^{N-1}Y_n=\sum_{n=1}^{\tau_\nu(j)-1}y_n+\sum_{n=\tau_\nu(j)}^{\tau_{\nu+1}(j)-1}y_n+\sum_{n=\tau_{\nu+1}(j)}^{\tau_N(i)-1}y_n. \tag{8}$$

We have by (6)

$$\left\{\begin{aligned}\boldsymbol{E}(|W|^p)&\leqq\boldsymbol{E}\left\{\left(\sum_{n=1}^{N-1}|Y_n|\right)^p\right\}\leqq\boldsymbol{E}\left\{\left(\sum_{n=1}^{N'}|Y_n|+\sum_{n=N'+1}^{N'+N''}|Y_n|\right)^p\right\}\\&\leqq 2^p\boldsymbol{E}\left\{\left(\sum_{n=1}^{N'}|Y_n|\right)^p\right\}+2^p\boldsymbol{E}\left\{\left(\sum_{n=N'+1}^{N'+N''}|Y_n|\right)^p\right\}.\end{aligned}\right. \tag{9}$$

Applying the second part of Theorem 3, we see that the last two expectations are equal so that

$$\boldsymbol{E}(|W|^p)\leqq 2^{p+1}\boldsymbol{E}\left\{\left(\sum_{n=1}^{N'}|Y_n|\right)^p\right\}.$$

If $p > 1$, using HÖLDER's inequality we obtain

$$(10)\quad \left\{\begin{aligned} \boldsymbol{E}\left\{\left(\sum_{n=1}^{N'}|Y_n|\right)^p\right\} &\leqq \boldsymbol{E}\left\{(N')^{p-1}\sum_{n=1}^{N'}|Y_n|^p\right\}\\ &= \sum_{m=1}^{\infty}\boldsymbol{P}\{N'(\omega)=m\}m^{p-1}\boldsymbol{E}\left\{\sum_{n=1}^{m}|Y_n|^p \,\Big|\, N'(\omega)=m\right\}. \end{aligned}\right.$$

We have by the definition of N':

$$\boldsymbol{E}\{|Y_n|^p|N'(\omega)=m\}=\boldsymbol{E}\{|Y_n|^p|x_s(\omega)\neq j \text{ for all } s\in I_n(i;\omega)\},\ 1\leqq n<m;$$
$$\boldsymbol{E}\{|Y_m|^p|N'(\omega)=m\}=\boldsymbol{E}\{|Y_m|^p|x_s(\omega)=j \text{ for some } s\in I_m(i;\omega)\}.$$

It follows from Theorem 13.5 that the right members above are independent of n and m respectively; hence we may denote them by E_1 and E_2. The hypothesis $\boldsymbol{E}(|Y_n|^p)<\infty$ for every n implies that $E_1<\infty$ and $E_2<\infty$. Substituting into (10) we obtain by (6) and (7),

$$\boldsymbol{E}\left\{\left(\sum_{n=1}^{N'}|Y_n|\right)^p\right\} \leqq \sum_{m=1}^{\infty}\boldsymbol{P}\{N'(\omega)=m\}m^{p-1}\{(m-1)E_1+E_2\}<\infty.$$

If $0\leqq p\leqq 1$, this inequality is valid if we replace m^{p-1} in it by 1. Therefore we have for $p\geqq 0$:

$$(11)\qquad \boldsymbol{E}(|W|^p)<\infty.$$

Now the three sums in the last member of (8) are independent random variables by the Corollary to Theorem 13.3; and it is well known that if W_1 and W_2 are independent random variables then $\boldsymbol{E}(|W_1+W_2|^p)<\infty$ implies $\boldsymbol{E}(|W_1|^p)<\infty$. Thus we conclude from (8) and (11) that

$$\boldsymbol{E}\left\{\left|\sum_{n=\tau_\nu(j)}^{\tau_{\nu+1}(j)-1} y_n\right|^p\right\}<\infty.$$

Since $\tau_\nu(j)$ is optional, the last expectation is equal to $\boldsymbol{E}\{|Y(j)|^p\}$ by the second part of Theorem 3, although ν is a random variable. Since j is arbitrary this proves the theorem. ▯

In case $f=1$, Theorem 4 reduces to a part of the assertion in Theorem 11.1 for an empty H in a recurrent class.

We now proceed to derive explicit formulas for the first two moments of $Y(i)$ under further assumptions. Let us first note that every f with domain $\boldsymbol{I}$ may be written as a linear combination of the Kronecker delta functions as follows:

$$(12)\qquad f(\cdot)=\sum_{j\in\boldsymbol{I}} f(j)\,\delta_j(\cdot)$$

where $\delta_j(i)=\delta_{ij}$ has already been used in the preceding proof.

Theorem 5. we have

$$E\{Y(i)\} = \sum_j {}_ip^*_{ij}\ (j) \tag{13}$$

provided that the series converges absolutely.

Proof. Taking expectations in (5) and using (12), we have

$$E\{Y(i)\} = f(i) + E\left\{\sum_{n=1}^{\infty} \hat{f}(\xi_n)\right\} = f(i) + E\left\{\sum_{n=1}^{\infty}\sum_j \hat{f}(j)\,\delta_j(\xi_n)\right\}.$$

If we interchange the expectation with the double summation, we have

$$E\{Y(i)\} = f(i) + \sum_{n=1}^{\infty}\sum_j \hat{f}(j)\,E\{\delta_j(\xi_n)\}. \tag{14}$$

Now if $j \neq i$ it follows from Theorem 13.5 that

$$E\{\delta_j(\xi_n)\} = P\{\xi_n(\omega) = j\} = {}_ip^{(n)}_{ij}.$$

Substituting this into (14) and interchanging the order of summation, we obtain

$$E\{Y(i)\} = f(i) + \sum_j \hat{f}(j) \sum_{n=1}^{\infty} {}_ip^{(n)}_{ij} = f(i) + \sum_j \hat{f}(j)\,{}_ip^*_{ij}.$$

This is equivalent to (13) since ${}_ip^*_{ii} = 1$. Now both interchanges are justified if the three operations E, $\sum_n$ and $\sum_j$ performed in some order on the absolute value of $\hat{f}(j)\,\delta_j(\xi_n)$ yield a finite result. Hence the absolute convergence of the series in (13) is a sufficient condition. ▯

We recall the notation $e_{ij} = {}_ip^*_{ij}$ introduced in §9. For any function f from $\boldsymbol{I}$ to $(-\infty, \infty)$ we introduce the new notation:

$$\mathfrak{J}_i(f) = \sum_j e_{ij} f(j); \tag{15}$$

and we say that $\mathfrak{J}_i(f)$ is *finite* in case $\mathfrak{J}_i(|f|) < \infty$, namely that the series in (15) converges absolutely. Thus (13) may be restated as follows:

$$E\{Y(i)\} = \mathfrak{J}_i(f)$$

provided that $\mathfrak{J}_i(f)$ is finite.

Theorem 6. If $\mathfrak{J}_i(f)$ is $\begin{Bmatrix}\text{finite}\\ \text{not zero}\end{Bmatrix}$ for any i, then it is $\begin{Bmatrix}\text{finite}\\ \text{not zero}\end{Bmatrix}$ for every i. Moreover if $\mathfrak{J}_i(f)$ and $\mathfrak{J}_i(g)$ are both finite then $\mathfrak{J}_i(f)/\mathfrak{J}_i(g)$ is independent of i.

Proof. Let $h \in \boldsymbol{I}$, then $e_{hj} = e_{hi}e_{ij}$ by Corollary 1 to Theorem 9.5. Hence we have

$$\mathfrak{J}_h(f) = e_{hi}\mathfrak{J}_i(f)$$

which proves the first assertion since $0 < e_{hi} < \infty$. Next, it follows from the above equation that

$$\frac{\mathfrak{J}_h(f)}{\mathfrak{J}_h(g)} = \frac{\mathfrak{J}_i(f)}{\mathfrak{J}_i(g)};$$

hence the second assertion. ▯

If the recurrent class $\boldsymbol{I}$ is positive, then $e_{ij} = \pi_j/\pi_i$ by a remark following Theorem 9.5; consequently

$$(16) \qquad \pi_i \mathfrak{J}_i(f) = \sum_j \pi_j f(j).$$

In this case $\{\pi_i, i \in \boldsymbol{I}\}$ is the stationary absolute distribution (see §7). If this is the initial distribution then the M.C. $\{x_n, n \geqq 0\}$ is stationary by the Corollary to Theorem 7.2, and we have

$$\boldsymbol{E}\{f(x_n)\} = \pi_i \mathfrak{J}_i(f), \qquad n \geqq 0.$$

Turning to the second moment of $Y(i)$, we shall be more brief since the method is the same.

Theorem 7. We have

$$(17) \qquad \boldsymbol{E}\{Y(i)^2\} = \sum_j {}_ip_{ij}^* f^2(j) + 2 \sum_j {}_ip_{ij}^* f(j) \sum_{k \neq i} {}_ip_{jk}^* f(k)$$

provided that both series converge absolutely.

Proof. As in the proof of Theorem 5 we have

$$\left(\sum_{n=\tau_1(i)}^{\tau_2(i)-1} y_n\right)^2 = f^2(i) + 2f(i) \sum_{n=1}^{\infty} \hat{f}(\xi_n) + \sum_{n=1}^{\infty} \hat{f}(\xi_n) + 2 \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \hat{f}(\xi_n) \hat{f}(\xi_{n+m}).$$

It follows from Theorem 13.5 that

$$\begin{aligned} \boldsymbol{E}\left\{\sum_{n=1}^{\infty} \hat{f}(\xi_n) \sum_{m=1}^{\infty} \hat{f}(\xi_{n+m})\right\} &= \sum_{n=1}^{\infty} \sum_j {}_ip_{ij}^{(n)} \hat{f}(j) \sum_{m=1}^{\infty} \sum_k {}_ip_{jk}^{(m)} \hat{f}(k) \\ &= \sum_j {}_ip_{ij}^* \hat{f}(j) \sum_k {}_ip_{jk}^* \hat{f}(k) \end{aligned}$$

provided that the last series converges absolutely. Hence we obtain

$$\boldsymbol{E}\{Y(i)^2\} = f^2(i) + 2f(i) \sum_j {}_ip_{ij}^* \hat{f}(j) + \sum_j {}_ip_{ij}^* \hat{f}^2(j) + 2 \sum_j {}_ip_{ij}^* \hat{f}(j) \sum_k {}_ip_{jk}^* \hat{f}(k)$$

which reduces to (17). ▯

Corollary. If $\boldsymbol{I}$ is a positive class, we have

$$\boldsymbol{E}\{Y(i)^2\} = m_{ii} \sum_j \frac{f^2(j)}{m_{jj}} + 2m_{ii} \sum_j \frac{f(j)}{m_{jj}} \sum_{k \neq i} \frac{m_{ji} + m_{ik} - m_{jk}}{m_{kk}} f(k)$$

provided that the series converge absolutely. In particular,

$$m_{ii}^{(2)} = m_{ii}\left(2\sum_j \frac{m_{ji}}{m_{jj}} - 1\right).$$

Proof. The first part follows from Corollary 2 to Theorem 11.3. Setting $f = 1$ and using Theorem 9.6, we have

$$\sum_{k \neq i} {}_ip_{jk}^{*} = m_{ji} - {}_ip_{ji}^{*} = m_{ji} - 1$$

whence the second part. ▯

Higher moments of $Y(i)$ can be evaluated in the same way but we shall not need them.

Before we turn to the last term $Y''(n)$ in the dissection formula, let us introduce a new notation:

$$U_\nu(i;\omega) = \sum_{n \in I_\nu(i;\,\omega)} |y_n(\omega)|, \qquad \nu \geqq 1. \tag{18}$$

Thus U_ν is to the functional $|f|$ as Y_ν is to f. The random variables $\{U_\nu, \nu \geqq 1\}$ are therefore independent, identically distributed with mean

$$\boldsymbol{E}\{U_\nu(i)\} = \mathfrak{I}_i(|f|). \tag{19}$$

We shall omit the subscript ν in an expression like $\boldsymbol{E}\{U_\nu(i)\}$ since it does not depend on ν, just as we have done with $\boldsymbol{E}\{Y_\nu(i)\}$.

Theorem 8. If $\boldsymbol{I}$ is a positive class, then $Y''(n)$ is bounded in probability, namely for any $\varphi(n)$ such that $\lim_{n\to\infty} \varphi(n) = \infty$, we have

$$\operatorname*{plim}_{n\to\infty} \frac{Y''(n)}{\varphi(n)} = 0 \tag{20}$$

where plim denotes the limit is probability. If $\boldsymbol{I}$ is a recurrent class and $\mathfrak{I}_i(|f|) < \infty$, then

$$\lim_{n\to\infty} \boldsymbol{E}\left\{\left|\frac{Y''(n)}{n}\right|\right\} = 0. \tag{21}$$

Proof. By the definition of $Y''(n)$ we have for every $u > 0$,

$$\boldsymbol{P}\{|Y''(n,\omega)| > u\} \leqq \boldsymbol{P}\left\{\max_{0\leqq s\leqq t}\left|\sum_{r=\tau_{l(n,\omega)}}^{\tau_{l(n,\omega)}+s} y_r(\omega)\right| > u\right\} + \boldsymbol{P}\{n - \tau_{l(n,\omega)}(\omega) > t\}.$$

The first term on the right tends to 0 as $u \to \infty$, for each fixed t. The second term tends to 0 uniformly in n as $t \to \infty$, by Theorem 14.2. This proves (20).

To prove (21) we need the following lemma:

Lemma. If $\eta_1, \ldots, \eta_n$ are identically distributed random variables with a finite mean, then

$$\lim_{n\to\infty} \frac{1}{n} \boldsymbol{E}\Big\{\max_{1\leqq\nu\leqq n} |\eta_\nu|\Big\} = 0.$$

Proof of the Lemma. Let the common distribution of the $|\eta_\nu|$ be F. We have

$$\boldsymbol{P}\Big\{\max_{1\leqq\nu\leqq n} |\eta_\nu| > k\Big\} \leqq n[1-F(k)].$$

It follows that

$$\sum_{k=0}^{\infty} \frac{1}{n} \boldsymbol{P}\Big\{\max_{1\leqq\nu\leqq n} |\eta_\nu| > k\Big\} \leqq \sum_{k=0}^{\infty} [1-F(k)] < \infty .$$

Thus the series converges dominatedly. Upon letting $n\to\infty$ each term tends to zero and the lemma is proved.

Now by definition we have, omitting i from the notation,

$$(22) \qquad |Y''(n,\omega)| \leqq U_{l(n,\omega)}(\omega).$$

It is important to remark that the random variable $U_{l(n)}$ need not have the same distribution as U_1. However, it is true that $1 \leqq l(n,\omega) \leqq n$. (If $l(n,\omega)=0$, then $Y''(n,\omega)=0$.) Hence we have

$$|Y''(n,\omega)| \leqq \max_{1\leqq\nu\leqq n} U_\nu(\omega).$$

Since $\boldsymbol{E}(U) < \infty$, (21) follows from the Lemma. ∎

Notes. Theorem 1 was obtained by the author in his lectures of 1953—54. The nondiscrete analogue has since been given by E. B. DYNKIN [1]; see also LAMPERTI [1]. The first assertion in Theorem 2 is given by KOLMOGOROV [3]. The basic idea embodied in Theorem 3 is due to DOEBLIN [4]. The special case where $f=1$ is implied in KOLMOGOROV [3], but the extension to $\{Y_\nu(i)\}$ must be considered as a brilliant stroke. Some will undoubtedly regard the truth of Theorem 3 as obvious, once it is enunciated, but this only illustrates the general lack of concern for rigor in probability theory.

The essential idea of the proof of Theorem 4 is due to J. WOLFOWITZ (oral communication). Most of the rest of the section is given in CHUNG [2] and [3], except the Lemma at the end.

§ 15. Ergodic theorems

We are now ready to resume the study of S_n. The first term $Y'(n)$ in the dissection formula (14.3) is simple; the second is the sum of a random number $l(n)-1$ of independent, identically distributed random variables Y_ν. There is dependence between $l(n)$ and the Y_ν's that has to be circumvented or analyzed. The third term $Y''(n)$, treated in Theorem 14.8, may at first sight appear quite similar to the first, but in reality it is somewhat more troublesome due to the greater dependence on n. In spite of these foreseen difficulties it is expected that the classical limit theorems for sums of independent identically distributed random variables should apply to S_n under suitable conditions. We proceed to investigate this. The following notation will be used in the sequel:

$$\text{Plim} = \text{limit with probability one;}$$

$$\text{plim} = \text{limit in probability;}$$

similarly for $\underline{\text{Plim}}$ and $\overline{\text{Plim}}$.

We recall that $\boldsymbol{I}$ is a recurrent class. Let f ang g be two functions from $\boldsymbol{I}$ to $(-\infty, \infty)$. In view of Theorem 14.6 we may drop the subscript i in $\mathfrak{I}_i(f)$ wherever immaterial.

Theorem 1. If $\mathfrak{I}(f)$ and $\mathfrak{I}(g)$ are both finite and not both zero, then we have

$$\underset{n\to\infty}{\text{Plim}} \frac{\sum_{s=0}^{n} f(x_s)}{\sum_{s=0}^{n} g(x_s)} = \frac{\mathfrak{I}(f)}{\mathfrak{I}(g)}.$$

Proof. Suppose that $f \geqq 0$; then it follows from (14.1) and (14.3) that

$$\sum_{\nu=1}^{l(n)-1} Y_\nu \leqq S_n \leqq Y'(n) + \sum_{\nu=1}^{l(n)} Y_\nu. \tag{1}$$

Since $\boldsymbol{I}$ is recurrent, $\underset{n\to\infty}{\text{Plim}}\, l(n) = +\infty$; it follows from (14.4) that

$$\underset{n\to\infty}{\text{Plim}} \frac{Y'(n)}{l(n)} = 0. \tag{2}$$

Applying the strong law of large numbers to the sequence $\{Y_\nu, \nu \geqq 1\}$, we have

$$\underset{l\to\infty}{\text{Plim}} \frac{1}{l} \sum_{\nu=1}^{l-1} Y_\nu = \boldsymbol{E}(Y). \tag{3}$$

By the nature of the convergence the above relation remains true if l is a random variable that tends to ∞ with probability one. Hence we may substitute $l(n)$ for l and we obtain, using (1) and (2) and Theorem 14.5,

$$\underset{n\to\infty}{\text{Plim}} \frac{S_n}{l(n)} = \boldsymbol{E}(Y) = \mathfrak{I}_i(f). \tag{4}$$

In the same way it is proved that the above relation holds if $f \leqq 0$. Hence it holds for a general f, provided $\mathfrak{I}_i(|f|) < \infty$. We have thus proved that

$$\underset{n\to\infty}{\mathrm{Plim}}\frac{1}{l(n)}\sum_{s=0}^{n} f(x_s)=\mathfrak{J}_i(f).$$

If $\mathfrak{J}(|g|)<\infty$ a similar relation holds with g replacing f. The theorem then follows by division, whereupon $l(n)$ cancels out. ▯

Theorem 1 may be used to reduce the study of a general functional f to a specific one g. For this purpose a good choice is $g(\cdot)=\delta_i(\cdot)$ for some i. The corresponding $S_n(\omega)$ is the number of values of s, $0\leqq s\leqq n$, for which $x_s(\omega)=i$. This will be denoted by $A_n(i;\omega)$ and the random variable by $A_n(i)$. Some of the known limit theorems concerning $A_n(i)$ can thus be immediately extended to S_n. Instead of going into these technical applications we state the following corollaries to Theorem 1.

Corollary. 1. For any j and k we have

$$\underset{n\to\infty}{\mathrm{Plim}}\frac{A_n(j)}{A_n(k)}=\frac{e_{ij}}{e_{ik}}.$$

Corollary 2. Let $\{x_n, n\geqq 0\}$ be a chain with independent and stationary increments whose state space forms a recurrent class. If

$$\sum_j |f(j)|<\infty, \quad \sum_j |g(j)|<\infty$$

then

$$\underset{n\to\infty}{\mathrm{Plim}}\frac{\sum_{s=0}^{n} f(x_s)}{\sum_{s=0}^{n} g(x_s)}=\frac{\sum_j f(j)}{\sum_j g(j)}.$$

In particular

$$\underset{n\to\infty}{\mathrm{Plim}}\frac{A_n(j)}{A_n(k)}=1.$$

This follows from Corollary 2 to Theorem 9.5. According to § 2 we may write $x_n=\sum_{s=0}^{n}\xi_s$ where $\{\xi_s, s\geqq 0\}$ is a sequence of independent random variables all of which except possibly ξ_0 have the same distribution. If all ξ_s are integer-valued with a common distribution whose mean is zero, then the M.C. is recurrent by a result of CHUNG and FUCHS [1].

We now specialize to a positive class.

Theorem 2. If I is a positive class, then

$$\underset{n\to\infty}{\mathrm{Plim}}\frac{1}{n}\sum_{s=0}^{n} f(x_s)=\sum_j \pi_j f(j) \tag{5}$$

provided that the series converges absolutely.

Proof. In a positive class $\mathfrak{J}_i(f)$ is given by (14.16); in particular $\mathfrak{J}_i(1)=\pi_i^{-1}$. Taking $g=1$ in Theorem 1 we obtain (5). ▯

Corollary 1. Let $l(n)$ and $\tau_{l(n)}$ be as in (14.1), then

$$\text{(6)} \qquad \operatorname*{Plim}_{n\to\infty} \frac{l(n)}{n} = \pi_i, \qquad \operatorname*{Plim}_{n\to\infty} \frac{\tau_{l(n)}}{n} = 1.$$

Proof. To prove the first assertion we need only take $f(\cdot) = \delta_i(\cdot)$ in Theorem 2. To prove the second assertion we write

$$\frac{\tau_{l(n)}}{n} = \frac{\tau_1 + \sum\limits_{\nu=1}^{l(n)-1} \varrho_\nu}{n} = \frac{\tau_1}{n} + \frac{l(n)-1}{n}\,\frac{1}{l(n)-1} \sum_{\nu=1}^{l(n)-1} \varrho_\nu.$$

Since τ_1 is a finite random variable, $\operatorname*{Plim}\limits_{n\to\infty} \tau_1 = 0$. By Theorem 14.3, the ϱ_ν are independent, identically distributed random variables with mean π_i^{-1}. Applying the strong law of large numbers and using also the first assertion we obtain

$$\operatorname*{Plim}_{n\to\infty} \frac{\tau_{l(n)}}{n} = 0 + \pi_i \frac{1}{\pi_i} = 1. \quad \square$$

Corollary 2. In a positive class, if $\mathfrak{I}(|f|) < \infty$, then

$$\operatorname*{Plim}_{n\to\infty} \frac{Y''(n)}{n} = 0.$$

This follows from the dissection formula and (2) to (5) above. Alternatively we have

$$\operatorname*{Plim}_{n\to\infty} \frac{U_{l(n,\omega)}(\omega)}{l(n,\omega)} = 0$$

as a consequence of the strong law of large numbers applied to the sequence $\{U_\nu\}$. Since $0 \leqq l(n,\omega) \leqq n$, Corollary 2 follows from (14.22). $\square$

Theorem 3. If $\boldsymbol{I}$ is a positive class, $\mathfrak{I}(|f|) < \infty$, and the initial distribution of the M.C. has finite support, then

$$\lim_{n\to\infty} \boldsymbol{E}\left\{\left|\frac{1}{n}\sum_{s=0}^{n} f(x_s) - \sum_j \pi_j\, f(\,)\right|\right\} = \lim_{n\to\infty} \boldsymbol{E}\left\{\left|\frac{S_n - \boldsymbol{E}(S_n)}{n}\right|\right\} = 0.$$

Proof. We recall the dissection formula. The hypothesis $\mathfrak{I}(|f|) < \infty$ means $\boldsymbol{E}(U) < \infty$ by (14.19) and consequently for every h,

$$\text{(7)} \qquad {}_i f^*_{ih}\, \boldsymbol{E}\{|Y'(n)|\,|\,x_0(\omega) = h\} \leqq \boldsymbol{E}(U) < \infty.$$

In what follows an integral over the set $\{\omega : l(n,\omega) = m\}$, e.g., will be denoted by $\int\limits_{\{l(n,\omega)=m\}} \ldots \boldsymbol{P}(d\omega)$. Since $l(n,\omega) \leqq n$ for every ω,

$$\begin{aligned}
\boldsymbol{E}\left\{\sum_{\nu=1}^{l(n)-1} Y_\nu\right\} &= \sum_{m=1}^{n} \int\limits_{\{l(n,\omega)=m\}} \sum_{\nu=1}^{m-1} Y_\nu(\omega)\,\boldsymbol{P}(d\omega) = \sum_{\nu=1}^{n-1} \int\limits_{\{l(n,\omega)>\nu\}} Y_\nu(\omega)\,\boldsymbol{P}(d\omega) \\
&= \sum_{\nu=1}^{n-1} \Big\{\boldsymbol{E}(Y_\nu) - \int\limits_{\{l(n,\omega)\leqq\nu\}} Y_\nu(\omega)\,\boldsymbol{P}(d\omega)\Big\}.
\end{aligned}$$

The last summation may be replaced by $\sum\limits_{\nu=1}^{n}$. Now it is clear that

$$\{\omega: l(n,\omega)<\nu\{=\{\omega: n<\tau_\nu(\omega)\}\in\mathscr{F}\{x_s,\, s\leqq\tau_\nu\}$$

while for every real c,

$$\{\omega: Y_\nu(\omega)=c\}\in\mathscr{F}\{x_s,\, s\geqq\tau_\nu\}.$$

Hence Y_ν is independent of the indicator of the set $\{\omega: l(n,\omega)<\nu\}$ by the Corollary to Theorem 13.3 and so

$$\int\limits_{\{l(n,\,\omega)<\nu\}} Y_\nu(\omega)\boldsymbol{P}(d\omega)=\boldsymbol{P}\{l(n,\omega)<\nu\}\boldsymbol{E}(Y_\nu).$$

We have therefore

$$\text{(8)}\quad \left\{\begin{aligned} \boldsymbol{E}\Big\{\sum_{\nu=1}^{l(n)-1} Y_\nu\Big\} &= \sum_{\nu=1}^{n}\Big[\boldsymbol{E}(Y)-\boldsymbol{P}\{l(n,\omega)<\nu\}\boldsymbol{E}(Y)-\int\limits_{\{l(n,\omega)=\nu\}} Y_\nu(\omega)\boldsymbol{P}(d\omega)\Big]\\ &=\boldsymbol{E}(Y)\sum_{\nu=1}^{n}\boldsymbol{P}\{l(n,\omega)\geqq\nu\}-\sum_{\nu=1}^{n}\int\limits_{\{l(n,\,\omega)=\nu\}} Y_\nu(\omega)\boldsymbol{P}(d\omega)\\ &=\boldsymbol{E}(Y)\boldsymbol{E}(l(n))-\sum_{\nu=1}^{n}\int\limits_{\{l(n,\,\omega)=\nu\}} Y_\nu(\omega)\boldsymbol{P}(d\omega). \end{aligned}\right.$$

In order to estimate the last term above we write

$$\begin{aligned}\Big|\sum_{\nu=1}^{n}\int\limits_{\{l(n,\,\omega)=\nu\}} Y_\nu(\omega)\boldsymbol{P}(d\omega)\Big| &\leqq\sum_{\nu=1}^{n}\int\limits_{\{l(n,\,\omega)=\nu\}}\max_{1\leqq\nu\leqq n}|Y_\nu(\omega)|\,\boldsymbol{P}(d\omega)\\ &=\boldsymbol{E}\Big\{\max_{1\leqq\nu\leqq n}|Y_\nu|\Big\}.\end{aligned}$$

Applying the Lemma in § 14 we obtain from (8),

$$\lim_{n\to\infty}\frac{1}{n}\boldsymbol{E}\Big\{\sum_{\nu=0}^{l(n)-1} Y_\nu\Big\}=\boldsymbol{E}(Y)\lim_{n\to\infty}\frac{\boldsymbol{E}\{l(n)\}}{n}=\pi_i\boldsymbol{E}(Y)$$

by Corollary 1 to Theorem 2 and bounded convergence. It follows from this, (7) and (14.21) that

$$\lim_{n\to\infty}\boldsymbol{E}\Big\{\frac{S_n}{n}\,\Big|\,x_0(\omega)=h\Big\}=\pi_i\boldsymbol{E}(Y)=\sum_j\pi_j f(j).$$

This being true for every h the same relation holds under any initial distribution with a finite support. We have therefore proved

$$\text{(9)}\qquad \lim_{n\to\infty}\frac{1}{n}\boldsymbol{E}(S_n)=\sum_j\pi_j f(j).$$

The following lemma on integration is well known (see e.g. Doob [4; p. 629]).

Lemma. Let $\{\eta_n,\, n\geqq 1\}$ be a sequence of nonnegative random variables such that $\operatorname*{plim}\limits_{n\to\infty}\eta_n=\eta$. Suppose that all $\boldsymbol{E}(\eta_n)$ and $\boldsymbol{E}(\eta)$ are finite. Then

$$\lim_{n\to\infty}\boldsymbol{E}(\eta_n)=\boldsymbol{E}(\eta)$$

if and only if the η_n are uniformly integrable, viz.

$$\lim_{N\to\infty} \int_{\{\eta_n(\omega)>N\}} \eta_n(\omega)\,\boldsymbol{P}(d\omega)=0$$

uniformly in n.

Applying the result (9) to the functional $|f|$ of $\{x_n\}$, we obtain

$$\lim_{n\to\infty} \boldsymbol{E}\left\{\frac{1}{n}\sum_{s=0}^{n}|f(x_s)|\right\}=\sum_j \pi_j |f(j)|.$$

It follows from Theorem 2 and one half of the Lemma applied to $|f|$ that the sequence $\left\{n^{-1}\sum_{s=0}^{n}|f(x_s)|,\ n\geqq 1\right\}$, and *a fortiori* the sequence $\{n^{-1}|S_n|,\ n\geqq 1\}$ is uniformly integrable. Hence we have from Theorem 2 and the other half of the Lemma,

$$\lim_{n\to\infty} \boldsymbol{E}\left\{\left|\frac{1}{n}\sum_{s=0}^{n} f(x_s)-\sum_j \pi_j f(j)\right|\right\}=\boldsymbol{E}\left\{\lim_{n\to\infty}\left|\frac{1}{n}\sum_{s=0}^{n} f(x_s)-\sum_j \pi_j f(j)\right|\right\}=0.$$

This and (9) prove the theorem. ▯

It is natural to ask, in view of (3), whether the existence of $\mu_i=\boldsymbol{E}\{Y(i)\}$ is sufficient for the validity of (5) if its right member is replaced by $\boldsymbol{E}\{Y(i)\}$. This condition is equivalent to

$$(10) \qquad \boldsymbol{E}\{|Y_\nu(i)|\}=\boldsymbol{E}\left\{\left|\sum_{n\in I_\nu(i)} f(x_n)\right|\right\}<\infty$$

while the condition imposed in Theorem 2 is equivalent to

$$(11) \qquad \boldsymbol{E}\{U_\nu(i)\}=\boldsymbol{E}\left\{\sum_{n\in I_\nu(i)} |f(x_n)|\right\}<\infty.$$

The answer to the question is negative, the reason being that the condition (10) is not sufficient to ensure that $\operatorname*{Plim}_{n\to\infty} Y''(n)/n=0$ although (11) is (Corollary 2 to Theorem 2). However, the weaker condition (10) does imply a weaker version of (5), namely Theorem 4 below. Before we prove this let us write the dissection formula in another form which will be used henceforth.

Let $i\in\boldsymbol{I}$ where $\boldsymbol{I}$ is a positive class. We put, recalling the $\varrho_\nu(i)$ defined at the beginning of § 14,

$$M_i=\pi_i\,\mu_i=\frac{\boldsymbol{E}\{Y(i)\}}{\boldsymbol{E}\{\varrho(i)\}};$$

$$\bar{f}=f-M_i;$$

$$z_n(\omega)=y_n(\omega)-M_i=\bar{f}[x_n(\omega)];$$

$$Z_\nu(i;\omega)=Y_\nu(i;\omega)-M_i\varrho_\nu(i;\omega)=\sum_{n\in I_\nu(i;\,\omega)} \bar{f}[x_n(\omega)]=\sum_{n\in I_\nu(i;\,\omega)} z_n(\omega).$$

(Note that $M_i=\pi_i\mathfrak{I}_i(f)$ if the latter is finite.)

We have then

$$\boldsymbol{E}\{Z_\nu(i)\} = \boldsymbol{E}\{Y_\nu(i)\} - M_i \boldsymbol{E}\{\varrho_\nu(i)\} = 0.$$

The dissection formula can be rewritten as follows, omitting i from the notation:

$$S_n - \Theta_n M(\tau_{l(n)} - \tau_1) = Y'(n) + \sum_{\nu=1}^{l(n)-1} Z_\nu + Y''(n) \tag{12}$$

where $Y'(n)$ and $Y''(n)$ are the same as in (14.3) and $\Theta_n = \Theta(i; n, \omega)$ is the indicator of $\{\omega: l(i; n, \omega) \geqq 1\}$. The random variables Z_ν are now independent and identically distributed with mean zero.

Theorem 4. If $\boldsymbol{I}$ is a positive class and μ_i exists, then

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{s=0}^{n} f(x_s) = M_i. \tag{13}$$

Proof. From (14.4) we have

$$\operatorname*{plim}_{n\to\infty} \frac{Y'(n)}{n} = 0.$$

By Theorem 14.8 we have

$$\operatorname*{plim}_{n\to\infty} \frac{Y''(n)}{n} = 0.$$

It remains to consider

$$\frac{1}{n} \sum_{\nu=1}^{l(n)-1} Z_\nu = \frac{l(n)-1}{n} \, \frac{1}{l(n)-1} \sum_{\nu=1}^{l(n)-1} Z_\nu.$$

Applying the *strong* law of large numbers to the sequence $\{Z_\nu, \nu \geqq 1\}$, we obtain

$$\operatorname*{Plim}_{n\to\infty} \frac{1}{l(n)} \sum_{\nu=1}^{l(n)-1} Z_\nu = \boldsymbol{E}(Z) = 0.$$

Since $\operatorname*{Plim}_{n\to\infty} l(n) = \infty$ and $l(n) \leqq n$ it follows from the last two equations that

$$\operatorname*{Plim}_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{l(n)-1} Z_\nu = 0,$$

and *a fortiori* this is true if Plim is replaced by plim. Substituting into (12) we obtain

$$\operatorname*{plim}_{n\to\infty} \frac{S_n}{n} = M_i \operatorname*{plim}_{n\to\infty} \frac{\tau_{l(n)}}{n} = M_i$$

by the Corollary to Theorem 14.2. This proves the theorem. ▯

Corollary. If $\boldsymbol{I}$ is a positive class and μ_i exists for some i, then all $\pi_i \mu_i$ have a common value.

This strengthens Theorem 14.4. We shall see in the next section a similar phenomenon for the second moment (Corollary 2 to Theorem 16.1).

We give two examples to show, on the one hand that the condition in (10) is not sufficient, and on the other hand that the condition (11) is not necessary for the *existence* of the limit in (5) or (9).

Example 1. Let $\boldsymbol{I}$ be the set of nonnegative integers. Let $p_0=1$,

$$p_{0,2i-1}=\frac{6}{\pi^2 i^2}, \quad p_{2i-1,2i}=p_{2i,0}=1, \quad i\geqq 1.$$

All states form a positive class of period 3 and it is easy to verify by Theorem 7.1 that

$$m_{00}=3; \quad m_{2i-1,2i-1}=m_{2i,2i}=\frac{\pi^2 i^2}{2}, \quad i\geqq 1.$$

Let $f(0)=0$, $f(2i-1)=i$, $f(2i)=-i$ for $i\geqq 1$. We have therefore

$$Y_\nu(0;\omega)=\sum_{n\in I_\nu(0;\omega)} y_n(\omega)=0+i_\nu(\omega)-i_\nu(\omega)=0,$$

so that (10) is trivialy satisfied with $\mu_0=0$. Hence Theorem 4 is applicable and (13) holds with $M_0=0$. On the other hand, we have

$$\sum_i \pi_i |f(i)| = \sum_i \frac{4}{\pi^2 i} = \infty,$$

so that (11) is not satisfied. It is clear that for every $A>0$,

$$\boldsymbol{P}\{y_{3n+1}(\omega)\geqq A n\}=\sum_{i\geqq A n}\frac{6}{\pi^2 i^2}\sim\frac{6}{A\pi^2 n}.$$

Furthermore the events $\{\omega: y_{3n+1}(\omega)\geqq A n\}$, $n\geqq 1$, are independent since $x_{3n}(\omega)=0$ with probability one. Hence by the Borel-Cantelli lemma, we have

$$\boldsymbol{P}\{y_{3n+1}(\omega)\geqq A n \quad \text{i.o.}\}=1.$$

This and the fact $S_{3n}(\omega)=0$ together imply that

$$0=\underline{\operatorname{Plim}}_{n\to\infty}\frac{S_n}{n}<\overline{\operatorname{Plim}}_{n\to\infty}\frac{S_n}{n}=\infty.$$

Consequently the limit in (5) does not exist, nor does that in (9) since $\boldsymbol{E}(S_{3n})=0$, $\boldsymbol{E}(S_{3n+1})=\infty$.

Example 2. Let $\boldsymbol{I}$ consist of 0 and ordered pairs (r, s) of integers such that $r\geqq 1$, $1\leqq s\leqq 2^r$. Let $p_0=1$,

$$p_{00}=2^{-1}, \quad p_{0,(r,1)}=3^{-r}, \quad r\geqq 1;$$

$$p_{(r,s),(r,s+1)}=1, \quad p_{(r,2^r),0}=1, \quad r\geqq 1, \quad 1\leqq s\leqq 2^r-1.$$

It is easily seen that

$$m_{00}=\frac{1}{2}+\sum_{r=1}^{\infty}\frac{2^r+1}{3}=3,$$

$$m_{(r,s),(r,s)}=3^{r+1}, \quad r\geqq 1, \quad 1\leqq s\leqq 2^r.$$

Let $f(0)=0$, and

$$f((r,s))=\begin{cases}+\left(\frac{3}{2}\right)^r & \text{if } s \text{ is odd;}\\ -\left(\frac{3}{2}\right)^r & \text{if } s \text{ is even.}\end{cases}$$

Then we have

$$\sum_j \pi_j |f(j)|=\sum_{r=1}^{\infty}\frac{2^r}{3^{r+1}}\left(\frac{3}{2}\right)^r=\infty;$$

on the other hand it is clear that $\mu_i=0$ for every $i\in \boldsymbol{I}$. Now it is easy to see that for every $n\geqq 0$,

$$\text{(14)} \qquad \boldsymbol{P}\left\{|S_n(\omega)|=\left(\frac{3}{2}\right)^r\right\}\leqq\frac{1}{3^r}, \qquad r\geqq 1.$$

It follows that for every $\lambda>(\lg 3-\lg 2)/\lg 3$,

$$\sum_n \boldsymbol{P}\{|S_n(\omega)|>n^{\lambda}\}<\infty,$$

and consequently by the Borel-Cantelli lemma

$$\operatorname*{Plim}_{n\to\infty}\frac{1}{n^{\lambda}}S_n=0.$$

Thus the limit in (5) is 0. Furthermore it follows from (14) that $\boldsymbol{E}(S_n)$ is bounded in n; hence the limit in (9) is also 0.

Notes. Theorem 1 is given in CHUNG [3]. HARRIS and ROBBINS [1] proved it for a general state space using HOPF's ergodic theorem. Corollary 1 was proved before Theorem 1 by LÉVY [1] and HARRIS [1]. Theorem 2 can be derived without pain from G. D. BIRKHOFF's ergodic theorem, where the needed metric transitivity can be proved by martingale theory, as suggested by DOOB (in a letter). For our purposes our proofs are by far the simplest.

For the kind of application of Theorem 1 alluded to after its proof see the Notes on § 16. For generalization of ratio ergodic theoremes as well as ratio limit theorems (§ 9) to general state space, see JAIN [1], KRENGEL [1]. In the chain case the latter will allow ratios of the form $\sum_{n=1}^{N} p^{(n)}(i,A)/\sum_{n=1}^{N} p^{(n)}(j,B)$ where e.g. $p^{(n)}(i,A)=\sum_{k\in \boldsymbol{A}} p_{ik}^{(n)}$.

Theorem 3 corresponds of course to the L_1-convergence version of the ergodic theorem; for the L_2-convergence version see § 16. Unlike the other results here, these two do not extend to an arbitrary initial distribution because the $\boldsymbol{E}$'s in question need not then be finite under our conditions. Theorem 4 is in CHUNG [3], but the proof there is inaccurate since we need the strong, rather than the weak, law of large numbers — a rather unexpected twist of fortune.

An open problem is to find necessary and sufficient conditions for the ergodic theorems given here. The two examples show that "deterministic circuits" must be excluded in order to reach a satisfactory result. It may be objected that these circuits are rather artificial and we may combine e.g. the two states $2i-1$ and $2i$ in Example 1 into one without any loss of information on the transition of the M.C. Nevertheless we can modify these circuits slightly to destroy the complete determinism while still keeping the requisite properties to serve as counter-examples. For instance we may let $p_{2i,2i-1}=p_{2i,0}=\frac{1}{2}$ in Example 1.

§ 16. Further limit theorems

In this section we give several more limit theorems about S_n including the central limit theorem and the law of the iterated logarithm. The state space $\boldsymbol{I}$ will now be assumed to be positive class, in fact beginning with Theorem 3 below the stronger assumption that $m_{ii}^{(2)}<\infty$ for some and hence all i will be made.

Let us rewrite (15.12) as

$$S_n-\Theta_n M n=Y'(n)+\sum_{\nu=1}^{l(n)-1} Z_\nu+Y''(n)-\Theta_n M(n-\tau_{n(l)}+\tau_1) \tag{1}$$

where $M=\pi_i\mu_i$ is independent of i by the Corollary to Theorem 15.4. Let

$$\sigma_i^2=\boldsymbol{E}\{Z_\nu^2(i)\}$$

and

$$B=B_i=\pi_i\sigma_i^2.$$

Note that σ_i^2 is not the variance of $Y_\nu(i)$. Applying Theorem 14.4 to the functional $\bar{f}=f-M$ we see that if $\sigma_i^2<\infty$ for some $i\in\boldsymbol{I}$ then $\sigma_i^2<\infty$ for all $i\in\boldsymbol{I}$.

Theorem 1. If $0<\sigma_i^2<\infty$, then S_n is asymptotically normally distributed with mean Mn and variance Bn; viz. for every real u,

$$\lim_{n\to\infty}\boldsymbol{P}\left\{\frac{S_n(\omega)-Mn}{\sqrt{Bn}}\leq u\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-v^2/2}\,dv=\Phi(u).$$

Proof. Dividing (1) through by $\sqrt{Bn}$ and letting $n\to\infty$, we see that the first, third, and fourth terms on the right all tend to zero in probability by (14.4) and (14.20), which latter is applied once as given and

Univ. of Arizona Library

once with $f=1$. (For the second application we may also use Theorem 14.2.) Thus it remains only to prove that

$$\lim_{n\to\infty} \boldsymbol{P}\left\{\frac{1}{\sqrt{Bn}} \sum_{\nu=1}^{l(n,\omega)-1} Z_\nu(\omega) \leqq u\right\} = \Phi(u).$$

Let $0<\varepsilon<1$ and

$$(2) \qquad n'=[\pi_i(1-\varepsilon^3)\,n], \qquad n^*=[\pi_i\, n]+1, \qquad n''=[\pi_i(1+\varepsilon^3)\,n]-1,$$

where the square bracket denotes the integral part. It follows from Corollary 1 to Theorem 15.2 that there exists a number $n_0(\varepsilon)$ such that the ω-set

$$\Lambda=\{\omega: n'<l(n,\omega)-1\leqq n'' \text{ for all } n>n_0(\varepsilon)\}$$

has probability $>1-\varepsilon$. Choose $n_0(\varepsilon)$ so large that $n^*<n''$ if $n>n_0(\varepsilon)$. If $\omega\in\Lambda$ then for all $n>n_0(\varepsilon)$ we have

$$\left|\sum_{\nu=1}^{l(n,\omega)-1} Z_\nu(\omega)-\sum_{\nu=1}^{n^*} Z_\nu(\omega)\right| \leqq 2 \max_{n'<\nu'\leqq n''}\left|\sum_{\nu=n'+1}^{\nu'} Z_\nu(\omega)\right|.$$

By the well known KOLMOGOROV inequality (see e.g. FELLER [3, p. 220]) we have

$$(3) \qquad \boldsymbol{P}\left\{\max_{n'<\nu'\leqq n''}\left|\sum_{\nu=n'+1}^{\nu'} Z_\nu(\omega)\right| > 2^{-1}\varepsilon\,\sigma_i\sqrt{n^*}\right\} \leqq \frac{4\,(n''-n')\,\sigma_i^2}{\varepsilon^2\,\sigma_i^2\, n^*} \leqq 8\varepsilon.$$

Therefore we have

$$\boldsymbol{P}\left\{\left|\sum_{\nu=1}^{l(n,\omega)-1} Z_\nu(\omega)-\sum_{\nu=1}^{n^*} Z_\nu(\omega)\right| > \varepsilon\,\sigma_i\sqrt{n^*}\right\} \leqq \boldsymbol{P}(\Omega-\Lambda)+8\varepsilon \leqq 9\varepsilon.$$

Since ε is arbitrary this means

$$\operatorname*{plim}_{n\to\infty}\left\{\frac{1}{\sigma_i\sqrt{n^*}}\sum_{\nu=1}^{l(n,\omega)-1} Z_\nu(\omega)-\frac{1}{\sigma_i\sqrt{n^*}}\sum_{\nu=1}^{n^*} Z_\nu(\omega)\right\}=0.$$

The classical central limit theorem applied to the sequence $\{Z_\nu, \nu\geqq 1\}$ asserts that

$$\lim_{n\to\infty} \boldsymbol{P}\left\{\frac{1}{\sigma_i\sqrt{n^*}}\sum_{\nu=1}^{n^*} Z_\nu(\omega)\leqq u\right\}=\Phi(u).$$

It follows from the last two equations that

$$\lim_{n\to\infty} \boldsymbol{P}\left\{\frac{1}{\sigma_i\sqrt{n^*}}\sum_{\nu=1}^{l(n,\omega)-1} Z_\nu(\omega)\leqq u\right\}=\Phi(u).$$

This is equivalent to the desired result since

$$\lim_{n\to\infty}\frac{\sigma_i\sqrt{n^*}}{\sqrt{Bn}}=1. \quad \square$$

Corollary 1. If $\sigma_i^2=0$ then S_n-Mn bounded in probability.

Proof. In this case $\boldsymbol{P}\{Z_\nu(\omega)=0\}=1$ for every ν and the other terms in (1) are all bounded in probability by previous results. ▯

Corollary 2. All $\pi_i\sigma_i^2$ have a common value, 0 and ∞ included.

Proof. This justifies our notation $B=B_i$. If $0<\sigma_i^2<\infty$ then Theorem 1 holds; hence $\sigma_j^2>0$ by Corollary 1. Moreover $\sigma_j^2<\infty$ by Theorem 14.4 as already noted above. Thus $0<\sigma_j^2<\infty$ and Theorem 1 holds now with j instead of i. It follows that $B_i=B_j$. If $\sigma_i^2=0$, then $\sigma_j^2<\infty$ by Theorem 14.4; thus $\sigma_j^2=0$ by what has just been proved. ▯

Theorem 2. If $\sigma_i^2=0$ for some i, then there exists a function ψ of two states such that

$$\sum_{n=m}^{m'-1} z_n=\psi(x_m, x_{m'}), \qquad 0\leqq m<m'<\infty, \tag{4}$$

with probability one.

Proof. The condition $\sigma_i^2=0$ is equivalent to

$$\sum_{n=m}^{m'-1} z_n(\omega)=0 \tag{5}$$

on any set $\{\omega: x_m(\omega)=x_{m'}(\omega)=i\}$, $0\leqq m<m'<\infty$. By Corollary 2 to Theorem 1, this is then true for every i; in other words the above equation holds for all ω for which $x_m(\omega)=x_{m'}(\omega)$. Thus (4) is established whenever $x_m(\omega)=x_{m'}(\omega)$ with $\psi(i,i)=0$ for every i. If the theorem is not true, then there exist two states i and j and two sets

$$\Lambda_1=\{\omega: x_m(\omega)=i;\ x_n(\omega)=i_n,\ m<n<m';\ x_{m'}(\omega)=j\}$$
$$\Lambda_2=\{\omega: x_s(\omega)=i;\ x_n(\omega)=j_n,\ s<n<s';\quad x_{s'}(\omega)=j\}$$

with $\boldsymbol{P}(\Lambda_1)>0$, $\boldsymbol{P}(\Lambda_2)>0$ such that (recalling that $z_n=\bar{f}(x_n)$)

$$\sum_{n=m+1}^{m'-1}\bar{f}(i_n)\neq\sum_{n=s+1}^{s'-1}\bar{f}(j_n).$$

Since $j\rightsquigarrow i$ there exist $t>0$ and $\{k_n,\ 0<n<t\}$ such that

$$\boldsymbol{P}\{x_n(\omega)=k_n,\ 0<n<t;\ x_t(\omega)=i\,|\,x_0(\omega)=j\}>0$$

(see the remarks after (4.1)). It follows that the two sets

$$\Lambda_1^*=\Lambda_1\cap\{\omega: x_{m'+n}(\omega)=k_n,\ 0<n<t;\ x_{m'+t}(\omega)=i\}$$
$$\Lambda_2^*=\Lambda_2\cap\{\omega: x_{s'+n}(\omega)=k_n,\ 0<n<t;\ x_{s'+t}(\omega)=i\}$$

both have positive probability; and if $\omega_1 \in \Lambda_1^*$, $\omega_2 \in \Lambda_2^*$ then

$$\sum_{n=m}^{m'+t-1} z_n(\omega_1) \neq \sum_{n=s}^{s'+t-1} z_n(\omega_2). \tag{6}$$

But $\Lambda_1^* \subset \{\omega : x_m(\omega) = x_{m'+t}(\omega) = i\}$, $\Lambda_2^* \subset \{\omega : x_s(\omega) = x_{s'+t}(\omega) = i\}$. Thus (5) and (6) are contradictory. This proves the theorem. ▯

The following set of conditions will be assumed in the remaining theorems:

$$\begin{cases} m_{ii}^{(2)} = \boldsymbol{E}\left\{\left(\sum_{n \in I_\nu(i;\,\omega)} 1\right)^2\right\} < \infty; \\ \boldsymbol{E}\{U_\nu^2(i)\} = \boldsymbol{E}\left\{\left[\sum_{n \in I_\nu(i;\,\omega)} |f[x_n(\omega)]|\right]^2\right\} < \infty. \end{cases} \tag{7}$$

Together they imply

$$\boldsymbol{E}\{Y_\nu^2(i)\} < \infty, \quad \boldsymbol{E}\{Z_\nu^2(i)\} < \infty.$$

Furthermore we know that the values of these expectations do not depend on ν and their finiteness does not depend on i.

Theorem 3. Under the assumptions in (7), if the initial distribution of the M.C. has finite support, we have

$$\lim_{n\to\infty} \frac{1}{n} \boldsymbol{E}\left\{\left[\sum_{s=0}^{n} f(x_s) - Mn\right]^2\right\} = \lim_{n\to\infty} \frac{1}{n} \boldsymbol{E}\{[S_n - \boldsymbol{E}(S_n)]^2\} = B.$$

Proof. It follows from (15.7) that we may suppose that $p_i = 1$. We first prove a lemma which sharpens (15.9) under the present conditions.

Lemma. We have

$$\boldsymbol{E}(S_n) = Mn + o(n^{\frac{1}{2}}).$$

Proof of the lemma. It follows from the proof of Theorem 15.3, in particular (15.8) that

$$\begin{aligned} |\boldsymbol{E}(S_n) - \boldsymbol{E}(Y)\boldsymbol{E}(l(n))| &\leqq \boldsymbol{E}\{|Y'(n)|\} + \boldsymbol{E}\left\{\max_{1\leqq\nu\leqq n}|Y_\nu|\right\} + \boldsymbol{E}\{|Y''(n)|\} \\ &\leqq O(1) + \boldsymbol{E}\left\{\max_{1\leqq\nu\leqq n}|Y_\nu|\right\} + \boldsymbol{E}\left\{\max_{1\leqq\nu\leqq n} U_\nu\right\}. \end{aligned}$$

By SCHWARZ's inequality and the Lemma in § 14 applied to the U_ν^2, we have

$$\boldsymbol{E}\left\{\max_{1\leqq\nu\leqq n}|Y_\nu|\right\} \leqq \boldsymbol{E}\left\{\max_{1\leqq\nu\leqq n} U_\nu\right\} \leqq \left(\boldsymbol{E}\left\{\max_{1\leqq\nu\leqq n} U_\nu^2\right\}\right)^{\frac{1}{2}} = o(n^{\frac{1}{2}}).$$

Next we have

$$\boldsymbol{E}(l(n)) = \sum_{\nu=0}^{n} p_{ii}^{(\nu)} = n\,\pi_i + O(1)$$

by Theorem 6.5. The lemma follows from these estimates.

We now write, using (1) and noting that $\Theta_n = 1$ for $n \geqq 1$ under $p_i = 1$,

$$(8) \quad S_n - \boldsymbol{E}(S_n) = \sum_{\nu=1}^{l(n)-1} Z_\nu + Y'(n) + Y''(n) - M(n - \tau_{l(n)} + \tau_1) + Mn - \boldsymbol{E}(S_n).$$

Now $\boldsymbol{E}\{Y'(n)^2\}$ and $\boldsymbol{E}\{Y''(n)^2\}$ are bounded in n by (7). Recalling (14.1) and the definition of the ϱ_ν in § 14, we have

$$\boldsymbol{E}\{[n - \tau_{l(n)}]^2\} \leqq \boldsymbol{E}\left\{\max_{1 \leqq \nu \leqq n} \varrho_\nu^2\right\} = o(n)$$

by the Lemma of § 14 applied to the ϱ_ν^2's, since $\boldsymbol{E}(\varrho_\nu^2) = m_{ii}^{(2)} < \infty$. Applying MINKOWSKI's inequality to (8) and using these estimates and the Lemma above, we obtain

$$\boldsymbol{E}\{[S_n - \boldsymbol{E}(S_n)]^2\} = \boldsymbol{E}\{[S_n - Mn]^2\} + o(n) = \boldsymbol{E}\left\{\left(\sum_{\nu=1}^{l(n)-1} Z_\nu\right)^2\right\} + o(n).$$

It remains to prove that

$$\lim_{n\to\infty} \frac{1}{n} \boldsymbol{E}\left\{\left(\sum_{\nu=1}^{l(n)-1} Z_\nu\right)^2\right\} = B$$

and this will be done in a similar manner as the proof of Theorem 15.3. We have

$$\boldsymbol{E}\left\{\left(\sum_{\nu=1}^{l(n)-1} Z_\nu\right)^2\right\} = \sum_{m=1}^{n} \int_{\{l(n,\,\omega)=m\}} \left\{\sum_{s=1}^{m-1} Z_s^2(\omega) + 2 \sum_{1 \leqq s < t < m} Z_s(\omega) Z_t(\omega)\right\} \boldsymbol{P}(d\omega).$$

The method used in the proof of Theorem 15.3 yields

$$(9) \quad \sum_{m=1}^{n} \int_{\{l(n,\,\omega)=m\}} \sum_{s=1}^{m-1} Z_s^2(\omega) \boldsymbol{P}(d\omega) = Bn + o(n).$$

Next we have since $\boldsymbol{E}(Z_s) = 0$,

$$\begin{aligned} \sum_{m=1}^{n} \sum_{1 \leqq s < t < m} \int_{\{l(n,\,\omega)=m\}} Z_s(\omega) Z_t(\omega) \boldsymbol{P}(d\omega) &= \sum_{1 \leqq s < t < n} \int_{\{l(n,\,\omega)>t\}} Z_s(\omega) Z_t(\omega) \boldsymbol{P}(d\omega) \\ &= -\sum_{1 \leqq s < t < n} \int_{\{l(n,\,\omega)\leqq t\}} Z_s(\omega) Z_t(\omega) \boldsymbol{P}(d\omega). \end{aligned}$$

For any real numbers a and b, we have if $s < t$,

$$\{\omega : l(n,\omega) < t;\, Z_s(\omega) = a\} \in \mathscr{F}\{x_r, r \leqq \tau_t\}$$

$$\{\omega : Z_t(\omega) = b\} \in \mathscr{F}\{x_r, r \geqq \tau_t\}.$$

It follows that

$$\int_{\{l(n,\,\omega)<t\}} Z_s(\omega) Z_t(\omega) \boldsymbol{P}(d\omega)) = \boldsymbol{E}(Z_t) \int_{\{l(n,\,\omega)<t\}} Z_s(\omega) \boldsymbol{P}(d\omega) = 0.$$

It remains only to consider

$$(10) \quad \sum_{t=1}^{n-1} \int_{\{l(n,\,\omega)=t\}} \left[\sum_{s=1}^{t-1} Z_s(\omega)\right] Z_t(\omega) \boldsymbol{P}(d\omega).$$

Clearly the square of this is bounded by

$$\left(\boldsymbol{E}\left\{\max_{1\leqq t\leqq n}\left|\sum_{s=1}^{t-1} Z_s\right| \max_{1\leqq t\leqq n}|Z_t|\right\}\right)^2 \leqq \boldsymbol{E}\left\{\max_{1\leqq t\leqq n}\left|\sum_{s=1}^{t-1} Z_s\right|^2\right\}\boldsymbol{E}\left\{\max_{1\leqq t\leqq n} Z_t^2\right\}.$$

We have

$$\boldsymbol{E}\left\{\max_{1\leqq t\leqq n}\left[\sum_{s=1}^{t-1} Z_s\right]^2\right\} \leqq 4\boldsymbol{E}\left\{\left(\sum_{s=1}^{n-1} Z_s\right)^2\right\} \leqq 4n\,\sigma_i^2$$

since the Z_s's are independent with mean zero, by an inequality of MARCINKIEWICZ and ZYGMUND (see DOOB [4; p. 317]); and the second factor is $o(n)$ by the Lemma in § 14. Consequently the quantity in (10) does not exceed $O(n^{\frac{1}{2}})\,o(n^{\frac{1}{2}})=o(n)$. This and (9) prove the theorem. ▯

Theorem 4. Under the assumptions in (7) we have, if $B>0$,

$$\lim_{n\to\infty}\boldsymbol{P}\left\{\max_{0\leqq\nu\leqq n}[S_\nu(\omega)-M\nu]\leqq u\sqrt{Bn}\right\}=\Psi(u)=\begin{cases}0 & \text{if } u<0,\\ \sqrt{\dfrac{2}{\pi}}\displaystyle\int_0^u e^{-v^2/2}\,dv & \text{if } u\geqq 0;\end{cases}$$

$$\lim_{n\to\infty}\boldsymbol{P}\left\{\max_{0\leqq\nu\leqq n}|S_\nu(\omega)-M\nu|\leqq u\sqrt{Bn}\right\}$$

$$=\frac{4}{\pi}\sum_{m=0}^{\infty}\frac{(-1)^m}{2m+1}\exp\left\{-\frac{(2m+1)^2\pi^2}{8u^2}\right\},\qquad u\geqq 0.$$

Proof. Let $\varepsilon>0$ be arbitrary. Writing $\sigma=\sigma_i$ and using the notation in (2) we define

$$\Lambda_0=\left\{\omega:\max_{1\leqq\ '\leqq n^*}\sum_{\nu=1}^{\nu'}Z_\nu(\omega)\leqq u\,\sigma\sqrt{n^*}\right\}$$

$$\Lambda_1=\{\omega: n'<l(n,\omega)-1\leqq n''\}$$

$$\Lambda_2=\left\{\omega:\max_{n'<\nu'\leqq n''}\left|\sum_{\nu=n'+1}^{\nu'}Z_\nu(\omega)\right|\leqq\varepsilon\,\sigma\sqrt{n^*}\right\}$$

$$\Lambda_3=\left\{\omega:\max_{1\leqq\nu\leqq n}|Y'(\nu,\omega)|\leqq\varepsilon\,\sigma\sqrt{n^*}\right\}$$

$$\Lambda_4=\left\{\omega:\max_{1\leqq\nu\leqq n}|Y''(\nu,\omega)|\leqq\varepsilon\,\sigma\sqrt{n^*}\right\}$$

$$\Lambda_5=\left\{\omega:\max_{1\leqq\nu\leqq n}M\big(\nu-\tau_{l(\nu,\omega)}(\omega)+\tau_1(\omega)\big)\leqq\varepsilon\,\sigma\sqrt{n^*}\right\}.$$

Recalling (1) above we see that if $\omega\in\bigcap_{s=1}^{5}\Lambda_s$, and n is so large that $n^*<n''$, then

$$\max_{1\leqq\nu\leqq n}[S_\nu(\omega)-M\nu]\leqq(u+4\varepsilon)\sigma\sqrt{n^*}.$$

On the other hand, if $\omega\in\Omega-\Lambda_0$ then there exists a positive integer $m(\omega)\leqq n^*$ such that

$$\sum_{\nu=1}^{m(\omega)} Z_\nu(\omega) > u\,\sigma\sqrt{n^*}.$$

Thus if $\omega\in(\Omega-\Lambda_0)\Lambda_2$, there exists $m'(\omega)\leqq n'$ such that

$$\sum_{\nu=1}^{m'(\omega)} Z_\nu(\omega) > (u-\varepsilon)\,\sigma\sqrt{n^*}.$$

Hence if $\omega\in(\Omega-\Lambda_0)\Lambda_1\Lambda_2$, then $l(n,\omega)-1>n'$ so that there exists $\tilde{n}(\omega)<n$ such that $l(\tilde{n}(\omega),\omega)-1=m'(\omega)$ and consequently

$$\max_{1\leqq\nu\leqq n}\sum_{t=1}^{l(\nu,\omega)-1} Z_t(\omega) > (u-\varepsilon)\,\sigma\sqrt{n^*}. \tag{11}$$

Finally if $\omega\in(\Omega-\Lambda_0)\bigcap_{s=1}^{5}\Lambda_s$ it follows from (1) and (11) that

$$\max_{1\leqq\nu\leqq n}[S_\nu(\omega)-M\nu] > (u-4\varepsilon)\,\sigma\sqrt{n^*}.$$

We have thus proved that

$$\boldsymbol{P}\Big(\bigcap_{s=1}^{5}\Lambda_s\Big)\leqq\boldsymbol{P}\Big\{\max_{1\leqq\nu\leqq n}[S_\nu(\omega)-M\nu]\leqq(u+4\varepsilon)\,\sigma\sqrt{n^*}\Big\}.$$

$$\boldsymbol{P}\Big(\bigcap_{s=1}^{5}\Lambda_s-\bigcap_{s=0}^{5}\Lambda_s\Big)\leqq\boldsymbol{P}\Big\{\max_{1\leqq\nu\leqq n}[S_\nu(\omega)-M\nu]>(u-4\varepsilon)\,\sigma\sqrt{n^*}\Big\}.$$

The last equation is equivalent to

$$1-\boldsymbol{P}\Big(\bigcap_{s=1}^{5}\Lambda_s\Big)+\boldsymbol{P}\Big(\bigcap_{s=0}^{5}\Lambda_s\Big)\geqq\boldsymbol{P}\Big\{\max_{1\leqq\nu\leqq n}[S_\nu(\omega)-M\nu]\leqq(u-4\varepsilon)\,\sigma\sqrt{n^*}\Big\}.$$

By a theorem due to ERDÖS and KAC [1], we have

$$\lim_{n\to\infty}\boldsymbol{P}(\Lambda_0)=\Psi(u).$$

We know that

$$\lim_{n\to\infty}\boldsymbol{P}(\Lambda_1)=1.$$

We have from (3)

$$\boldsymbol{P}(\Lambda_2)\geqq 1-2\varepsilon.$$

It is easy to see from (14.4) that

$$\lim_{n\to\infty}\boldsymbol{P}(\Lambda_3)=1.$$

Now we come to a new estimate. We have

$$\boldsymbol{P}(\Omega-\Lambda_4)=\boldsymbol{P}\Big\{\max_{\tau_{l(n,\omega)}<\nu\leqq n}\Big|\sum_{t=\tau_{l(n,\omega)}}^{\nu} y_t(\omega)\Big|>\varepsilon\,\sigma\sqrt{n^*}\Big\}$$

$$\leqq\boldsymbol{P}\Big\{\max_{1\leqq\nu\leqq n}U_\nu(\omega)>\varepsilon\,\sigma\sqrt{n^*}\Big\}$$

since

$$\max_{\tau_{l(n,\omega)}<\nu\leqq n}\left|\sum_{t=\tau_{l(n,\omega)}}^{\nu} y_t(\omega)\right| \leqq \sum_{t=\tau_{l(n,\omega)}}^{\tau_{l(n,\omega)+1}-1} |y_t(\omega)| = U_{l(n,\omega)}(\omega)$$

and $1\leqq l(\nu,\omega)\leqq n$. Consequently we have

$$\begin{aligned} \boldsymbol{P}(\Omega-\Lambda_4) &\leqq n\boldsymbol{P}\left\{U_1(\omega)>\varepsilon\,\sigma\sqrt{n^*}\right\} \\ &\leqq \frac{n}{\varepsilon^2\sigma^2 n^*}\int\limits_{\{U_1(\omega)>\varepsilon\sigma\sqrt{n^*}\}} U_1^2(\omega)\boldsymbol{P}(d\omega)\to 0 \end{aligned}$$

as $n\to\infty$, namely,

$$\lim_{n\to\infty}\boldsymbol{P}(\Lambda_4)=1.$$

Finally it follows from the same argument that leads to the above estimate but applied to the functional $f=1$ instead, that

$$\lim_{n\to\infty}\boldsymbol{P}(\Lambda_5)=1.$$

Collecting all the estimates we obtain

$$\Psi(u-4\varepsilon)-2\varepsilon\leqq \overline{\underline{\lim}}_{n\to\infty}\boldsymbol{P}\left\{\max_{1\leqq\nu\leqq n}[S_\nu(\omega)-M\nu]\leqq u\sigma\sqrt{n^*}\right\}\leqq\Psi(u+4\varepsilon)+2\varepsilon.$$

Since Ψ is continuous this proves the first assertion of the theorem. The second assertion is proved in the same way by merely changing the definition of Λ_0 and applying the corresponding theorem of Erdös and Kac. ▯

We now come to the law of the iterated logarithm.

Theorem 5. Under the assumptions in (7) we have, if $B>0$,

$$\overline{\mathrm{Plim}}_{n\to\infty}\frac{S_n-Mn}{\sqrt{2Bn\lg\lg n}}=1$$

and the corresponding $\underline{\mathrm{Plim}}$ is equal to -1.

Proof. According to the law of the iterated logarithm in the form given by Hartman and Wintner [1], we have

$$\overline{\mathrm{Plim}}_{n\to\infty}\frac{\sum_{\nu=1}^{l(n)-1} Z_\nu}{\sqrt{2\sigma^2 l(n)\lg\lg l(n)}}=1,$$

since $\mathrm{Plim}_{n\to\infty}\, l(n)=\infty$. We know $Y'(n)$ is finite with probability one. Applying the strong law of large numbers to the sequence $\{U_\nu^2,\nu\geqq 1\}$ we have

$$\mathrm{Plim}_{n\to\infty}\frac{\sum_{\nu=1}^{l(n)} U_\nu^2}{l(n)}=\boldsymbol{E}(U^2)<\infty.$$

It follows that

$$\operatorname*{Plim}_{n\to\infty} \frac{U^2_{l(n)}}{l(n)} = 0$$

and consequently by (14.22)

$$\operatorname*{Plim}_{n\to\infty} \frac{Y''(n)}{\sqrt{l(n)}} = 0.$$

The last result applied to the functional $f=1$ yields also

$$\operatorname*{Plim}_{n\to\infty} \frac{n-\tau_{l(n)}}{\sqrt{l(n)}} = 0.$$

Theorem 5 now follows from (1) and the above relations. ▯

Remark. *Both in Theorem 4 and in Theorem 5 we may replace* $S_\nu - M\nu$ *by* $S_\nu - \boldsymbol{E}(S_\nu)$ *and* Bn *by* $\boldsymbol{E}\{[S_n - \boldsymbol{E}(S_n)]^2\}$, on account of Theorem 3 and the lemma there.

Theorems 3, 4, and 5 are proved under the assumptions in (7). The condition that $\boldsymbol{E}\{U^2(i)\} < \infty$ implies that the evaluation of $\boldsymbol{E}\{Y^2(i)\}$ given in the Corollary to Theorem 14.7 is valid. That expression involves the "mixed moments" m_{ij} as well as the "pure moments" m_{jj}. Now under other restrictive hypothesis it has been shown (Doob [4]) that the condition

$$\sum_j \frac{|f(j)|^r}{m_{jj}} < \infty$$

for some $r>2$ or even for $r=2$ is sufficient for the validity of the central limit theorem. The following example shows that this is not so in general.

Example. Let $\boldsymbol{I}$ be the set of nonnegative integers. Let $\alpha_0=1$, $0<\alpha_n<1$ if $n\geqq 1$; $p_0=1$ and

$$p_{n0} = 1-\alpha_n, \quad p_{n,n+1} = \alpha_n, \quad n \geqq 0.$$

Put

$$\beta_0 = 1, \quad \beta_n = \prod_{\nu=0}^{n-1} \alpha_\nu.$$

Obviously all the states form a class of period 1. We have $f_{00}^{(1)}=0$ and

$$f_{00}^{(n)} = \beta_{n-1}(1-\alpha_{n-1}) = \beta_{n-1} - \beta_n, \quad n \geqq 2.$$

Hence the class is recurrent if and only if $\lim_{n\to\infty} \beta_n = 0$. Let c be such that $\sum_{n=2}^{\infty} c n^{-3} (\lg n)^{-2} = 1$. Take

$$\beta_n = \sum_{\nu=n+1}^{\infty} \frac{c}{\nu^3 \lg^2 \nu}, \quad n \geqq 1.$$

This defines α_n for $n \geqq 1$ and gives

$$f_{00}^{(n)} = \frac{c}{n^3 \lg^2 n}, \qquad n \geqq 2.$$

Thus $m_{00}^{(2)} < \infty$. We have furthermore $p_{ij}^{(n)} = p_{i,j-1}^{(n-1)} \alpha_{j-1}$ whence we obtain

$$\frac{1}{m_{jj}} = \frac{\alpha_{j-1}}{m_{j-1,j-1}}$$

and consequently

$$\frac{1}{m_{jj}} = \frac{\beta_j}{m_{00}} \sim \frac{c}{2 m_{00} j^2 \lg^2 j}, \qquad j \to \infty.$$

Moreover a direct computation yields

$$\begin{aligned} m_{j0} &= (1-\alpha_j) + 2\alpha_j (1-\alpha_{j+1}) + 3\alpha_j \alpha_{j+1} (1-\alpha_{j+2}) + \cdots \\ &= \frac{1}{\beta_j} \left(f_{00}^{(j+1)} + 2 f_{00}^{(j+2)} + 3 f_{00}^{(j+3)} + \cdots\right) \sim j, \qquad j \to \infty. \end{aligned}$$

We define the function f as follows: for a certain $r > 2$:

$$f(j) = j^{1/r} \qquad j \geqq 0.$$

In the following $c_1, \ldots, c_4$ denote positive constants. We have then

$$\sum_j \frac{f^r(j)}{m_{jj}} \leqq c_1 \sum_{j=2}^{\infty} \frac{j}{j^2 \lg^2 j} < \infty.$$

Take the fixed state i in §14 to be 0 and write $\varrho_\nu(\omega) = \varrho_\nu(0; \omega)$, $Y_\nu(\omega) = Y_\nu(0; \omega)$ etc. By the structure of the M.C. we have

$$\boldsymbol{P}\left\{Y_1(\omega) = \sum_{j=1}^{n-1} j^{1/r}\right\} = f_{00}^{(n)}.$$

Since $\sum_{j=1}^{n-1} j^{1/r} \sim \frac{r}{r+1} n^{(r+1)/r}$ $\left(= n(r), \text{say}\right)$ it follows that as $x \to \infty$,

$$\boldsymbol{P}\{Y_1(\omega) > x\} \sim \sum_{n(r) > x} f_{00}^{(n)} = \sum_{n > x(r)} \frac{c}{n^3 \lg^2 n} \sim \frac{c_2}{x^{\frac{2r}{r+1}} \lg^2 x}, \tag{12}$$

where $x(r) = [(r+1)\, x/r]^{r/(r+1)}$. On the other hand, we have

$$\left\{ \begin{aligned} \int_0^{x-} y^2\, d_y \boldsymbol{P}\{Y_1(\omega) \leqq y\} &\sim \sum_{n(r) < x} \left(\frac{r}{r+1} n^{\frac{r+1}{r}}\right)^2 f_{00}^{(n)} \\ &= c_3 \sum_{n < x(r)} \frac{n^{\frac{2(r+1)}{r}}}{n^3 \lg^2 n} \sim c_4 \frac{x^{\frac{2}{r+1}}}{\lg^2 x}. \end{aligned} \right. \tag{13}$$

Therefore

$$\lim_{x\to\infty} \frac{x^2 \boldsymbol{P}\{Y_1(\omega) > x\}}{\int\limits_0^{x-} y^2 \, d_y \, \boldsymbol{P}\{Y_1(\omega) \leqq y\}} = \frac{c_2}{c_4} > 0 .$$

Thus the necessary and sufficient condition for the validity of the central limit theorem for a sequence of independent, identically distributed random variables (see GNEDENKO-KOLMOGOROV [1, p. 172]) is not satisfied and we conclude that there do not exist two sequences $\{A_n\}$ and $\{B_n\}$ where $B_n > 0$ such that

$$\frac{1}{B_n}\left(\sum_{\nu=1}^{n} Y_\nu - A_n\right)$$

converges in distribution to Φ. We now proceed to show that the same is true if $\sum\limits_{\nu=1}^{n} Y_\nu$ is replaced by $\sum\limits_{\nu=1}^{l(n)-1} Y_\nu$ in the above. It will then follow that the same conclusion holds if $\sum\limits_{\nu=1}^{n} Y_\nu$ is replaced by S_n, on account of (14.3), (14.4), and (14.20). This means that the central limit theorem fails in every conceivable way for the sequence $\{S_n\}$.

As a consequence of the law of the iterated logarithm applied to the sequence $\{\varrho_\nu\}$ we have for every $\varepsilon > 0$,

$$\operatorname*{Plim}_{n\to\infty} \frac{\sum\limits_{\nu=1}^{n} \varrho_\nu - \dfrac{n}{\pi_0}}{n^{\frac{1}{2}+\varepsilon}} = 0 .$$

This is equivalent to the following assertion: given any $\varepsilon > 0$, there exists an $n_0(\varepsilon)$ and a set Λ_0 with $\boldsymbol{P}(\Lambda_0) > 1 - \varepsilon$ such that if $\omega \in \Lambda_0$,

$$|l(n, \omega) - \pi_0 n| < n^{\frac{1}{2}+\varepsilon} \quad \text{for all} \quad n > n_0(\varepsilon) .$$

Let

$$n' = [\pi_0 n - n^{\frac{1}{2}+\varepsilon}], \qquad n^* = [\pi_0 n], \qquad n'' = [\pi_0 n + n^{\frac{1}{2}+\varepsilon}] .$$

Then if $\omega \in \Lambda_0$, we have

$$\left| \sum_{\nu=1}^{l(n,\omega)-1} Y_\nu(\omega) - \sum_{\nu=1}^{n^*} Y_\nu(\omega) \right| \leqq \left| \sum_{\nu=n'}^{n''} Y_\nu(\omega) \right|$$

since each $Y_\nu \geqq 0$ with probability one. Let ϑ be a positive number to be specified later. We have by (12) and (13),

$$\sum_{\nu=n'}^{n''} \boldsymbol{P}\{Y_\nu(\omega) > n^\vartheta\} = o\left(n^{\frac{1}{2}+\varepsilon-\frac{2r\vartheta}{r+1}}\right);$$

$$\sum_{\nu=n'}^{n''} \int\limits_0^{n^\vartheta -} y^2 \, d_y \, \boldsymbol{P}\{Y_\nu(\omega) \leqq y\} = o\left(n^{\frac{1}{2}+\varepsilon+\frac{2\vartheta}{r+1}}\right).$$

Applying CHEBYSHEV'S inequality to Y_ν, $n' \leqq \nu \leqq n''$, truncated at n^ϑ, we obtain

$$P\left\{\left|\sum_{\nu=n'}^{n''} Y_\nu(\omega)\right| > n^{\frac{1-\varepsilon}{2}}\right\} \leqq o\left(n^{\frac{1}{2}+\varepsilon-\frac{2r\vartheta}{r+1}}\right) + o\left(n^{\frac{1}{2}+\varepsilon+\frac{2\vartheta}{r+1}-1+\varepsilon}\right).$$

Hence if we choose

$$\frac{(1+2\varepsilon)(r+1)}{4r} < \vartheta < \frac{(1-4\varepsilon)(r+1)}{4}$$

the above probability will be $o(1)$. Such a choice is possible whenever $r > 1$, by choosing ε sufficiently small. With such a choice we have then

$$(14) \quad \overline{\lim_{n\to\infty}}\, P\left\{\left|\sum_{\nu=1}^{l(n,\omega)-1} Y_\nu(\omega) - \sum_{\nu=1}^{n^*} Y_\nu(\omega)\right| > n^{\frac{1-\varepsilon}{2}}\right\} \leqq P(\Omega - \Lambda_0) \leqq \varepsilon.$$

Now suppose that

$$(15) \qquad \frac{1}{B_n}\left(\sum_{\nu=1}^{l(n)-1} Y_\nu - A_n\right)$$

tends in distribution to Φ as $n \to \infty$ and consider two cases.

Case 1. We have $\sqrt{n} = O(B_n)$. This implies by (14) that

$$\frac{1}{B_n}\left(\sum_{\nu=1}^{n^*} Y_\nu - A_n\right)$$

tends in distribution to Φ. Since $\lim_{n\to\infty} n^*/n = \pi_0 < 1$ it follows that there exist A'_n and $B'_n > 0$ such that $\frac{1}{B'_n}\left(\sum_{\nu=1}^{n} Y_\nu - A'_n\right)$ tends in distribution to Φ. This has been shown to be impossible.

Case 2. There exists a sequence $\{n_k\}$ such that $B_{n_k} = o(\sqrt{n_k})$. It follows then from (14) that

$$(16) \qquad \operatorname*{plim}_{k\to\infty} \frac{1}{\sqrt{n_k^*}}\left(\sum_{\nu=1}^{n_k^*} Y_\nu - A_{n_k}\right) = 0.$$

According to the necessary and sufficient conditions for the validity of the weak law of large numbers (see GNEDENKO-KOLMOGOROV [1; p. 134]) it is necessary that

$$n_k^*\{|Y_1(\omega) - m| \geqq \sqrt{n_k^*}\} = o(1)$$

where m is a median of Y_1. But the left member is asymptotically equal to

$$n_k^* \frac{c_2}{n_k^{*\,\frac{r}{r+1}} \lg^2 n_k^*}$$

by (12); consequently (16) is false and Case 2 is impossible.

We have therefore proved that in either case the sequence of random variables (15) does not tend in distribution to Φ and with this our goal is reached.

Notes. Theorems 1 and 2 are due to DOEBLIN [4] and [1]. Our proof of Theorem 1 is slightly simpler than DOEBLIN'S and, thanks to an observation due to D. G. KENDALL [3], does not require the condition $m_{ii}^{(2)} < \infty$.

Theorem 3 is new and is a sharp form of the L_2 version of the ergodic theorem. Theorems 4 and 5 are in CHUNG [3] but the proofs briefly indicated there are incorrect due to the fact that $U_{l(n)}$ need not have the same distribution as U_1. For the history leading to the counter-example at the end of the section see loc. cit.; it has now been strengthened to permit the most general normalization.

In the case of a finite number of states, material on the moments of S_n and the central limit theorem can be found in FRÉCHET [1], ROMANOVSKIĬ [1] and SARYMSAKOV [1]. The discussion in the last-mentioned book is however not always reliable (see the author's review in *Math. Reviews*, 16, p. 1179; in addition to the errors mentioned there, the proof of the law of the iterated logarithm on pp. 138—139 of the book is also inaccurate). For a local central limit theorem using an improvement of DOEBLIN'S method, see KOLMOGOROV [4]; the extension of this result to the denumerable case would be of interest.

It is clear that DOEBLIN'S method is also applicable to the case where the sequence $\{Y_\nu(i), i \geq 1\}$ belongs to the domain of attraction of a stable distribution. Indeed such a result is given in DOEBLIN [4], but the conditions imposed are such as to render it essentially a restatement of the known result for an independent sequence. The results of FELLER [2] for the case $f = \delta_i$ are in the same category but where they can be extended to a general f their interest increases from the point of view of M.C. theory. In some cases this extension can be made by a simple application of Theorem 15.1. For example, such an application to Theorem 7, part (1) of FELLER [2] yields a reasonable result concerning a limit stable distribution with index < 1 for S_n. (This result includes a recent one of DOBRUŠIN [1] as a special case.) Unfortunately the method is inapplicable to part (2) of FELLER'S theorem concerning a stable distribution of index > 1, owing to the circumstance that $\mathfrak{I}(f) = \mathfrak{I}(g) = 0$ for the appropriate f and g. This happens in general when the random variables $Y_\nu(i)$ have to be normalized by subtracting the mean, for then $\mathfrak{I}(\bar{f}) = 0$. Such is the case also with FELLER'S Theorem 5 which is the particular case of our Theorem 1 for $f = \delta_i$. Thus the result is valid, but the method of Theorem 15.1 fails. How to replace the latter theorem in the indeterminate case is an open question. On the other hand, FELLER'S method does not seem to yield DOEBLIN'S theorem.

§ 17. Almost closed and sojourn sets

In this last section of Part I we return to general considerations of the evolution of the M.C. The new results will concern primarily non-recurrent (essential or inessential) states. Logically speaking, this section may be placed immediately after § 4 as it is largely independent of the developments thereafter. Its postponement is partly due to the methodological difference and partly due to its being an introduction to an as yet unexplored part of the theory of M.C.

Let $\boldsymbol{I}^\infty$ be the space of *infinite state sequences*, namely sequences of elements of $\boldsymbol{I}$. A generic element of $\boldsymbol{I}^\infty$ may be denoted by $\boldsymbol{j}=\{j_0, j_1, \ldots\}$. Define a transformation T on $\boldsymbol{I}^\infty$, called the *shift*, by

$$T\{j_0, j_1, \ldots\}=\{j_1, j_2, \ldots\}.$$

Let f be a function with domain $\boldsymbol{I}^\infty$ and range the finite real line. The function f is said to be *invariant (under the shift)* iff $f(T\boldsymbol{j})=f(\boldsymbol{j})$ for every $\boldsymbol{j}\in\boldsymbol{I}^\infty$. Now consider the M.C. $\{x_n, n\geqq 0\}$ with state space $\boldsymbol{I}$ on the triple $(\Omega, \mathscr{F}, \boldsymbol{P})$. It is well known (see DOOB [4; p. 603]) that a function φ defined on Ω and measurable with respect to $\mathscr{F}\{x_n, n\geqq 0\}$ is a Baire function f of the sequence $\boldsymbol{x}=\{x_0, x_1, \ldots\}$:

$$\varphi(\omega)=f(x_0(\omega), x_1(\omega), \ldots), \qquad \omega\in\Omega. \tag{1}$$

For a given φ there may be more than one f satisfying (1); e.g., if $x_0\equiv j_0$ on Ω then the identically vanishing function φ may be represented in the form (1) by either the identically vanishing f or the function $f\equiv x_0-j_0$. The function φ is said to be *invariant* iff there exists an invariant Baire function f for which (1) holds. A set Λ in $\mathscr{F}\{x_n, n\geqq 0\}$ is said to be *invariant* iff its indicator φ_Λ is an invariant function and hence representable in the form (1) by an invariant Baire function f_Λ of $\boldsymbol{x}$. It follows that for every $n\geqq 0$,

$$\varphi_\Lambda(\omega)=f_\Lambda(T^n\boldsymbol{x})=f_\Lambda(x_n(\omega), x_{n+1}(\omega), \ldots). \tag{2}$$

The collection of all invariant sets is a nonempty Borel subfield of $\mathscr{F}$; this field will be called the *invariant field* and denoted by $\mathscr{G}$. It is not necessarily complete with respect to $\boldsymbol{P}$. The above definition may be relaxed by allowing an exceptional null set of ω in (1). Two functions on Ω are said to be *equivalent* iff their values coincide except on a null set; and two sets Λ_1 and Λ_2 in $\mathscr{F}$ are said to be *equivalent* iff their *symmetric difference* $(\Lambda_1-\Lambda_1\Lambda_1)\cup(\Lambda_2-\Lambda_1\Lambda_2)$ is a null set. We write $\Lambda_1\doteq\Lambda_2$ when Λ_1 and Λ_2 are equivalent, and we write $\Lambda_1 \dot{\subset} \Lambda_2$ when $\Lambda_1 - \Lambda_1\Lambda_2$ is a null set. When the equivalence classes of equivalent functions or sets are considered we describe the situation as usual by the phrase "modulo null sets". This leads to the relaxed definition of "invariance" just mentioned.

Let A be an arbitrary subset of $\boldsymbol{I}$ and define two sets of $\mathscr{G}$ as follows:

$$\overline{\mathscr{L}}(A) = \limsup_n \{\omega\colon x_n(\omega) \in A\}, \qquad \underline{\mathscr{L}}(A) = \liminf_n \{\omega\colon x_n(\omega) \in A\}.$$

Thus $\omega \in \overline{\mathscr{L}}(A)$ if and only if $x_n(\omega) \in A$ "infinitely often" and $\omega \in \underline{\mathscr{L}}(A)$ if and only if $x_n(\omega) \in A$ "ultimately". Let φ_n^A be the indicator of the set $\{\omega\colon x_n(\omega) \in A\}$. Then the indicators of $\overline{\mathscr{L}}(A)$ and $\underline{\mathscr{L}}(A)$ are respectively $\limsup_n \varphi_n^A$ and $\liminf_n \varphi_n^A$ both of which are clearly invariant Baire functions of $\boldsymbol{x}$. It follows that $\overline{\mathscr{L}}(A)$ and $\underline{\mathscr{L}}(A)$ are invariant sets, the first one having already been considered in §4. Obviously $\underline{\mathscr{L}}(A) \subset \overline{\mathscr{L}}(A)$; furthermore if we use the symbol "$'$" to denote complementation both in $\boldsymbol{I}$ and in Ω, we have $\underline{\mathscr{L}}(A) = (\overline{\mathscr{L}}(A'))'$. The set A will be called *transient* iff $\boldsymbol{P}(\overline{\mathscr{L}}(A)) = 0$; it is easily seen that this is the case if and only if $\boldsymbol{P}\{\overline{\mathscr{L}}(A) \mid x_n(\omega) = i\} = 0$ for every $n \geqq 0$ and $i \in A$. Thus if A consists of a single state, transience reduces to nonrecurrence (see the Corollary to Theorem 4.3). The set A will be called *almost closed* iff

$$\boldsymbol{P}(\overline{\mathscr{L}}(A)) = \boldsymbol{P}(\underline{\mathscr{L}}(A)) > 0. \tag{3}$$

A closed set in the sense of §3 is almost closed. If A is transient or almost closed then $\overline{\mathscr{L}}(A) \doteq \underline{\mathscr{L}}(A)$; in this case we shall write $\mathscr{L}(A)$ for any set which is equivalent to $\overline{\mathscr{L}}(A)$ or $\underline{\mathscr{L}}(A)$. In our previous notation $\lim_n \varphi_n^A$ then exists and is equivalent to the indicator of $\mathscr{L}(A)$. Thus to each almost closed set A there corresponds an invariant set which is equivalent to $\mathscr{L}(A)$. The following basic theorem establishes the converse and elaborates on this correspondence.

Theorem 1. To each invariant set Λ there corresponds a transient or almost closed set A such that $\Lambda \doteq \mathscr{L}(A)$, according as Λ is a null set or not. This correspondence is one-to-one modulo null sets on the one hand and transient sets on the other. Furthermore it is an isomorphism with respect to complementation, union and intersection.

Proof. If Λ is an invariant set, we have, using the notation in (2),

$$\begin{aligned}\boldsymbol{E}\{f_\Lambda(\boldsymbol{x}) \mid x_0 = i_0, \ldots, x_n = i_n\} &= \boldsymbol{E}\{f_\Lambda(T^n \boldsymbol{x}) \mid x_0 = i_0, \ldots, x_n = i_n\} \\ &= \boldsymbol{E}\{f_\Lambda(T^n \boldsymbol{x}) \mid x_n = i_n\} = \boldsymbol{E}\{f_\Lambda(\boldsymbol{x}) \mid x_0 = i_n\},\end{aligned}$$

where the first equation follows from (2), the second from the Markov property and the third from stationarity. This really means that there is a function ψ on $\boldsymbol{I}$ such that

$$\boldsymbol{E}\{f_\Lambda(x) \mid x_0, \ldots, x_n\} = \psi(x_n) \tag{4}$$

with probability one. The left member of (4), being the conditional expectation of a given bounded random variable relative to an increasing sequence of Borel fields, is a martingale. A theorem in martingale theory (Doob [4; p. 332]) asserts that there is a set Ω_0 with $\boldsymbol{P}(\Omega_0)=1$ such that we have on Ω_0,

$$\lim_{n\to\infty} \psi(x_n)=\boldsymbol{E}\{f_\Lambda(\boldsymbol{x})\,|\,x_n,\ 0\leqq n<\infty\}=f_\Lambda(\boldsymbol{x}).$$

Choose an arbitrary constant a, $0<a<1$, and define the set of states A as follows:

$$A=\{i: \psi(i)>a\}. \tag{5}$$

We are going to prove that

$$\overline{\mathscr{L}}(A)\doteq\Lambda\doteq\underline{\mathscr{L}}(A) \tag{6}$$

so that A is transient or almost closed according as Λ is or is not a null set, and in either case $\Lambda\doteq\mathscr{L}(A)$. To prove (6) we observe that if $\omega_0\in\Omega_0\Lambda$, then $f_\Lambda(\boldsymbol{x}(\omega_0))=1$; consequently $\psi(x_n(\omega_0))>a$ or $x_n(\omega_0)\in A$ for all sufficiently large n; hence $\omega_0\in\underline{\mathscr{L}}(A)$. On the other hand if $\omega_0\in\Omega_0-\Lambda$ then $f_\Lambda(\boldsymbol{x}(\omega_0))=0$; consequently $\psi(x_n(\omega_0))<a$ or $x_n(\omega_0)\notin A$ for all sufficiently large n; hence $\omega_0\notin\overline{\mathscr{L}}(A)$. We have therefore $\overline{\mathscr{L}}(A)\mathrel{\dot{\subset}}\Lambda\mathrel{\dot{\subset}}\underline{\mathscr{L}}(A)$, establishing (6).

Clearly for two almost closed sets A and B, $\mathscr{L}(A)\doteq\mathscr{L}(B)$ if and only if the symmetric difference of A and B is transient. Taking complements in (6) we obtain

$$\underline{\mathscr{L}}(A')\doteq\Lambda'\doteq\overline{\mathscr{L}}(A').$$

Next, let $\mathsf{M}\doteq\overline{\mathscr{L}}(B)\doteq\underline{\mathscr{L}}(B)$, then

$$\underline{\mathscr{L}}(A)\cup\underline{\mathscr{L}}(B)\subset\underline{\mathscr{L}}(A\cup B)\subset\overline{\mathscr{L}}(A\cup B)\subset\overline{\mathscr{L}}(A)\cup\overline{\mathscr{L}}(B).$$

The extreme terms are both equivalent to $\Lambda\cup\mathsf{M}$; hence the middle terms are both equivalent to the same. This establishes the asserted isomorphism. □

If Λ is an invariant set, then the value of $\boldsymbol{P}\{\Lambda\,|\,x_n=i\}$ does not depend on n whenever defined, as a consequence of (4), and we shall write $n=0$ there according to our notational convention made at the beginning of §4.

Corollary. If Λ is invariant, then $\Lambda\doteq\mathscr{L}(A)$ where

$$A=\{i: \boldsymbol{P}\{\Lambda\,|\,x_0(\omega)=i\}>a\}$$

and a is an arbitrary number satisfying $0<a<1$.

For an arbitrary set of states A let us write

$$p^{(n)}(i,A)=\boldsymbol{P}\{x_n(\omega)\in A\,|\,x_0(\omega)=i\}.$$

Clearly we have

$$(7) \qquad \boldsymbol{P}\{\underline{\mathscr{L}}(A)\,|\,x_0(\omega)=i\} \leqq \overline{\lim_{n\to\infty}}\, p^{(n)}(i, A) \leqq \boldsymbol{P}\{\overline{\mathscr{L}}(A)\,|\,x_0(\omega)=i\}.$$

Hence if A is almost closed then for any $m \geqq 0$ and $i \in \boldsymbol{I}$,

$$\boldsymbol{P}\{\mathscr{L}(A)\,|\,x_m(\omega)=i\} = \lim_{n\to\infty} p^{(n)}(i, A)$$

exists. (The converse is not true: take A to be a single nonrecurrent state.) Next, it follows from the second inequality in (7) that if A is transient then $\lim_{n\to\infty} p^{(n)}(i, A)=0$. (The converse is not true: take A to be a single null recurrent state.)

A set of states S is called a *sojourn set* iff $\boldsymbol{P}(\underline{\mathscr{L}}(S))>0$. This is equivalent to the apparently stronger condition that there exist $i\in S$ and $N\geqq 0$ such that

$$(8) \qquad \boldsymbol{P}\{x_n(\omega)\in S \text{ for all } n\geqq N\,|\,x_N(\omega)=i\}>0.$$

For it is a consequence of the definition that there exists an $N<\infty$ such that

$$0< \boldsymbol{P}\{x_n(\omega)\in S, n\geqq N\} = \sum_{k\in \boldsymbol{I}} \sum_{i\in S} p_k\, p_{ki}^{(N)}\, \boldsymbol{P}\{x_n(\omega)\in S, n\geqq N+1\,|\,x_N(\omega)=i\},$$

from which (8) follows.

Let S be a sojourn set and $0<a<1$; we put

$$S^{(a)} = \{i: \boldsymbol{P}\{\underline{\mathscr{L}}(S)\,|\,x_0(\omega)=i\}>a\}.$$

Theorem 2. If S is a sojourn set then for every a, $0<a<1$, the set $S\,S^{(a)}$ is almost closed, $S^{(a)}-S\,S^{(a)}$ is transient, and

$$\boldsymbol{P}\{\underline{\mathscr{L}}(S)\,|\,x_0(\omega)=i\} = \lim_{n\to\infty} p^{(n)}(i, S\,S^{(a)}).$$

Proof. The set $\underline{\mathscr{L}}(S)$ being invariant, $S^{(a)}$ is almost closed and $\underline{\mathscr{L}}(S) \doteq \mathscr{L}(S^{(a)})$ by the Corollary to Theorem 1. Consequently

$$\underline{\mathscr{L}}(S) \doteq \underline{\mathscr{L}}(S)\,\underline{\mathscr{L}}(S^{(a)}) = \underline{\mathscr{L}}(S\,S^{(a)})$$

and

$$\overline{\mathscr{L}}(S\,S^{(a)}) \subseteq \mathscr{L}(S^{(a)}) \doteq \underline{\mathscr{L}}(S) \doteq \underline{\mathscr{L}}(S\,S^{(a)}).$$

Hence $S\,S^{(a)}$ is almost closed. Moreover, $\mathscr{L}(S^{(a)})$ and $\mathscr{L}(S\,S^{(a)})$ are equivalent so that $S^{(a)}-S\,S^{(a)}$ is transient. The last assertion of the theorem now follows from (7). ▯

The set $S-S\,S^{(a)}$ need not be transient: let S be the union of a null recurrent state and a sojourn set consisting only of nonrecurrent states.

Theorem 3. If S is a sojourn set and $S\subset A$, then

$$\boldsymbol{P}\{\underline{\mathscr{L}}(A)\}=\boldsymbol{P}\{\underline{\mathscr{L}}(S\,S^{(a)})\}+\boldsymbol{P}\{\underline{\mathscr{L}}(A-S\,S^{(a)})\}.$$

Proof. In general we have if $B\subset A$,

$$\underline{\mathscr{L}}(B)\cup\underline{\mathscr{L}}(A-B)\subset\underline{\mathscr{L}}(A)\subset\underline{\mathscr{L}}(B)\cup\underline{\mathscr{L}}(A-B)\cup[\overline{\mathscr{L}}(B)\cap\overline{\mathscr{L}}(A-B)].$$

If $B=S\,S^{(a)}$ which is almost closed, then

$$\overline{\mathscr{L}}(S\,S^{(a)})\,\overline{\mathscr{L}}(A-S\,S^{(a)})\doteq\underline{\mathscr{L}}(S\,S^{(a)})\,\overline{\mathscr{L}}(A-S\,S^{(a)})=0.$$

The theorem follows since $\underline{\mathscr{L}}(B)$ and $\underline{\mathscr{L}}(A-B)$ are disjoint sets. ▯

Let $\mathfrak{A}$ be an arbitrary Borel subfield of $\mathscr{F}$. Relative to $(\Omega, \mathfrak{A}, \boldsymbol{P})$ a set Λ in $\mathfrak{A}$ is called *atomic* iff $\boldsymbol{P}(\Lambda)>0$ and Λ does not contain two disjoint subsets in $\mathfrak{A}$ of positive probability; it is called *completely nonatomic* iff $\boldsymbol{P}(\Lambda)>0$ and Λ does not contain any atomic subset. In the latter case for each given c, $0<c<\boldsymbol{P}(\Lambda)$, there exists a subset M in $\mathfrak{A}$ of Λ such that $\boldsymbol{P}(\mathsf{M})=c$. The following lemma is well known (sec e.g. Loève [1, p. 100]).

Lemma. Relative to $(\Omega, \mathfrak{A}, \boldsymbol{P})$ we have the following decomposition:

(9) $$\Omega=\Lambda_0\cup\Lambda_1\cup\Lambda_2\cup\cdots$$

where the Λ's are a finite or denumerable number of disjoint sets in $\mathfrak{A}$, at most one of which is completely nonatomic and the others are atomic. The decomposition is unique modulo null sets in $\mathfrak{A}$.

Let us apply this lemma to $\mathfrak{A}=\mathscr{G}$ and use the notation in (9). To each invariant set Λ_n of positive probability there corresponds an almost closed set of states A_n. In analogy with the definitions above an almost closed set will be called *atomic* iff it does not contain two disjoint almost closed sets, and *completely nonatomic* iff it does not contain any atomic almost closed set.

Theorem 4. We have the following decomposition of the state space:

(10) $$\boldsymbol{I}=A_0\cup A_1\cup A_2\cup\cdots$$

where the A's are a finite or denumerable number of disjoint almost closed sets, at most one of which is completely nonatomic and the others are atomic; and

(11) $$\sum_{n=0}^{\infty}\boldsymbol{P}\{\underline{\mathscr{L}}(A_n)\}=1.$$

The decomposition is unique modulo transient sets.

Proof. Let B_n correspond to Λ_n in the Lemma with $\mathfrak{A}=\mathscr{G}$, according to Theorem 1. Let $A_0=B_0$, $A_n=B_n-\bigcup_{\nu=0}^{n-1}B_\nu B_n$, $n\geqq 1$. It follows from the isomorphism of the correspondence that $B_\nu B_n$ is transient if $\nu \neq n$, hence $B_n - A_n$ is transient. Thus A_n as well as B_n is almost closed and $\mathscr{L}(A_n) \doteq \mathscr{L}(B_n) \doteq \Lambda_n$. The atomicity of the sets A_n and the uniqueness follow from the isomorphism and (11) follows from (9). The set $\boldsymbol{I} - \bigcup_{n=0}^{\infty} A_n$ is transient and may be absorbed in any A_n. □

Corollary. Each existing recurrent class may be taken as one of the atomic A_n's; each of the remaining A_n's including the completely nonatomic one, if present, contains only nonrecurrent states.

This follows from the fact, consequence of the Corollary to Theorem 4.5, that a recurrent class forms an atomic almost closed set.

Example 1. Let $\boldsymbol{I}$ be the set of nonnegative integers, $p_{01} = p_{02} = \frac{1}{2}$; $p_{2i-1,\,2i+1} = p_{2i,\,2i+2} = 1 - 1/(i+1)^2$, $p_{2i-1,\,0} = p_{2i,\,0} = 1/(i+1)^2$, $i \geqq 1$. Then $\boldsymbol{I}$ is a nonrecurrent class; the set of even integers and the set of odd integers are two disjoint atomic almost closed sets.

Example 2. Let $\boldsymbol{I}$ be the set of ordered pairs of nonnegative integers. $p_{(i,0),(i+1,0)} = 1 - 1/(i+2)^2$, $p_{(i,0),(i,1)} = 1/(i+2)^2$, $p_{(i,j),(i,j+1)} = 1$, $i \geqq 0$, $j \geqq 1$. All states are inessential. The atomic almost closed sets are $A_0 = \{(i,0),\ i \geqq 0\}$, $A_{i+1} = \{(i,j),\ j \geqq 1\}$, $i \geqq 0$; which yield the decomposition asserted in Theorem 4. Let $B_i = \{(i,j),\ j \geqq 0\}$, $i \geqq 0$; then the B_i's are disjoint atomic almost closed sets, but they do not yield the decomposition in the sense of Theorem 4 since (11) is not satisfied. There is no representative of A_0 among the B_i's.

Example 3. Let $\{\varepsilon_n,\ n \geqq 1\}$ be a sequence of independent and identically distributed random variables taking the values $+1$ and -1 with probability $\frac{1}{2}$ each; and let $x_n = 2^{-1} + \sum_{\nu=1}^{n} \varepsilon_\nu 2^{-\nu-1}$. Then $\{x_n,\ n \geqq 0\}$ is a M.C. with independent (but not stationary!) increments, of which the state space $\boldsymbol{I}$ is the set of numbers of the form $(2m+1)\,2^{-n}$, $1 \leqq 2m+1 < 2^n$, $n \geqq 1$. It is easy to see that $\boldsymbol{I}$ is a completely nonatomic (almost) closed set, since for each n there exist 2^n disjoint almost closed sets A_ν such that $\boldsymbol{P}\{\mathscr{L}(A_\nu)\} = 2^{-n}$.

We wind up with the discussion of the system of equations (7.9) which is "conjugate" to the determining system (7.1).

Theorem 5. To every bounded solution $\{u(i),\ i \in \boldsymbol{I}\}$ of the system of equations

$$u(i) = \sum_{j \in \boldsymbol{I}} p_{ij}\, u(j), \qquad i \in \boldsymbol{I} \tag{12}$$

there corresponds a bounded invariant function φ such that

$$u(x_n) = \boldsymbol{E}\{\varphi \mid x/_n\} \tag{13}$$

a.e., and conversely.

Proof. Suppose that (12) is true and consider the functional process $\{u(x_n), n \geqq 0\}$. We have

$$\boldsymbol{E}\{u(x_{n+1}) | x_0, \ldots, x_n\} = \boldsymbol{E}\{u(x_{n+1}) | x_n\} = \sum_j p_{x_n, j}\, u(j) = u(x_n).$$

Hence the sequence $\{u(x_n), \mathscr{F}_n\}$ where $\mathscr{F}_n = \mathscr{F}\{x_\nu, 0 \leqq \nu \leqq n\}$ is a martingale. Since u is bounded it follows from a martingale convergence theorem (Doob [4; p. 319]) that

$$\lim_{n \to \infty} u(x_n) = \varphi$$

exists and $\boldsymbol{E}\{\varphi | x_0\} = u(x_0)$ with probability one. Clearly φ is bounded and equivalent to an invariant function. Conversely, if (13) is true where φ is bounded and invariant, then by the Markov property:

$$u(x_n) = \boldsymbol{E}\{\varphi | x_0, \ldots, x_n\}.$$

Thus $\{u(x_n), \mathscr{F}_n\}$ is a martingale and (12) becomes the defining relation

$$u(x_n) = \boldsymbol{E}\{u(x_{n+1}) | x_n\}. \quad \square$$

Corollary. $\boldsymbol{I}$ consists of a single atomic almost closed set if and only if the only bounded solution of the system (12) is $u(\cdot) = \text{constant}$.

Notes. The main results, Theorems 1, 2, and 5, are due to Blackwell [3], to which we refer for application to chains with independent and stationary increments. For further results along these lines see Breiman [1]. Theorems 3 and 4 amount to new proofs of some auxiliary results in Feller [4], other related results can be derived by the same method. Apart from their intrinsic interest, these results will be useful in the future development of the continuous parameter theory; see Feller [5] and § 19 of Part II. The discussion of the transition probability $p^{(n)}(i, A)$ to a *set* becomes properly speaking a part of the general theory of Markov processes with an arbitrary, not necessarily denumerable, state space; see Chung [13].

It should be remarked that the notion of "shift" is used only in this section of the book. Its introduction requires either a specification of the sample space as discussed here, or a new axiomatic foundation as in Dynkin [2]. Our avoidance of it elsewhere in the book is done at some expense of extra work, for instance in § 15 of Part II. However, the novice is sternly warned against taking this shifty operation for granted.

For connections with discrete potential theory, see Doob [5], Hunt [1] and Neveu [4]. For applications to boundary theory see Chung [12].

Part II. Continuous parameter

§ 1. Transition matrix: basic properties

In Part II it will be convenient to begin with the analysis of a denumerable set of real valued functions, later (in § 4) to be identified as the transition probability functions of a continuous parameter Markov chain. Although the questions treated in the next three sections are of a purely analytic nature some of the methods used will be probabilistically inspired. Occasionally the results are given in a more general form than is required by later applications.

Henceforth $\boldsymbol{T}$ stands for the interval $[0, \infty)$, $\boldsymbol{T}^0$ for $(0, \infty)$; $\boldsymbol{I}$ for a denumerable set of indices, later to be specified to be the minimal state space of a Markov chain. Unless otherwise stated, the letters s, t, u, δ, ε denote positive real numbers, namely points of $\boldsymbol{T}^0$; i, j, k, l denote elements of $\boldsymbol{I}$; ν, m, n denote positive integers. An unspecified sum over the indices is over all $\boldsymbol{I}$.

A *transition matrix* is a finite or denumerable array of functions $(p_{ij}(\cdot))$ or more simply (p_{ij}), $i, j \in \boldsymbol{I}$, defined on $\boldsymbol{T}^0$ and satisfying the following three conditions: for every i, j and s, t,

(A) $$p_{ij}(t) \geqq 0;$$

(B) $$\sum_j p_{ij}(t) = 1;$$

(C) $$\sum_k p_{ik}(s) p_{kj}(t) = p_{ij}(s+t).$$

If we denote the matrix $(p_{ij}(t))$ by $P(t)$, then these properties can be stated as follows: each element of $P(t)$ is nonnegative; each row sum is equal to one; and the family $\{P(t), t \in \boldsymbol{T}^0\}$ is a semigroup with respect to the usual matrix multiplication. (Convergence of the row-column product in (C) in the infinite case is ensured by (A) and (B).) Conditions (A) and (B) together may be expressed by saying that for each t, the matrix $(p_{ij}(t))$ is *stochastic*. Condition (C) is often referred to as the *Chapman-Kolmogorov equation*.

In the sequel measurability on $\boldsymbol{T}$ or part of it means Lebesgue measurability, unless otherwise specified. This measure is denoted by μ and the phrases "almost all" and "almost everywhere", abbreviated to "a.a." and "a.e.", have the usual meaning.

An example of a transition matrix whose elements are not measurable will be given at the end of this monograph. Measurability of all the elements of the matrix has far-reaching consequences. We record this as a basic assumption:

(D) Every p_{ij} is a measurable function in $\boldsymbol{T}^0$.

A transition matrix which satisfies the condition (D) is said to be *measurable.*

Theorem 1. For any transition matrix (p_{ij}) and each fixed $h>0$ the sum

$$\sum_j |p_{ij}(t+h)-p_{ij}(t)|$$

is nonincreasing as t increases. If (p_{ij}) is measurable, then the above sum tends to zero uniformly in $t \geqq \delta > 0$ as $h \to 0$. In particular, each p_{ij} is uniformly continuous in $[\delta, \infty)$ for every $\delta > 0$.

Proof. We have if $0<s<t$,

$$\sum_j |p_{ij}(t+h)-p_{ij}(t)| = \sum_j \Big|\sum_k [p_{ik}(s+h)-p_{ik}(s)]\, p_{kj}(t-s)\Big|$$
$$\leqq \sum_k |p_{ik}(s+h)-p_{ik}(s)| \sum_j p_{kj}(t-s) = \sum_k |p_{ik}(s+h)-p_{ik}(s)|\,.$$

This proves the first assertion. Next if the matrix is measurable, then upon integrating we obtain, if $t \geqq \delta > 0$,

$$\sum_j |p_{ij}(t+h)-p_{ij}(t)| \leqq \sum_k \frac{1}{\delta} \int_0^\delta |p_{ik}(s+h)-p_{ik}(s)|\, ds.$$

Hence if $0 \leqq h \leqq \delta$ the second series is dominated by

$$\sum_k \frac{2}{\delta} \int_0^{2\delta} p_{ik}(s)\, ds$$

and consequently converges uniformly in $h \in [0, \delta]$. By a well known theorem (see e.g. TITCHMARSH [1; p. 377]) we have for each k,

$$\lim_{h \to 0} \int_0^\delta |p_{ik}(s+h)-p_{ik}(s)|\, ds = 0.$$

The second assertion of the theorem follows from this and the uniform convergence, and the third assertion is an immediate consequence. ▯

The most important part of Theorem 1 is: *every element p_{ij} of a measurable transition matrix is a continuous function in $\boldsymbol{T}^0$.*

In order to lead to a further essential hypothesis regarding the transition matrix, we study the behavior of $p_{ij}(t)$ as $t \to 0$ from the right. (Since $p_{ij}(t)$ is not defined for $t<0$ we shall simply write $t \to 0$ for such a limit.) We prove first an algebraic result.

Theorem 2. Let the matrix (u_{ij}) be given satisfying the following conditions for every i and j:

$$u_{ij} \geqq 0; \tag{1}$$

$$\sum_j u_{ij} \leqq 1; \tag{2}$$

$$u_{ij} \leqq \sum_k u_{ik} u_{kj}. \tag{3}$$

Then the index set $\boldsymbol{I}$ may be partitioned into disjoint sets $F, I, J, \ldots$ such that

$$u_{ij} = 0 \qquad \text{if } j \in F; \tag{4}$$

$$u_{ij} = \delta_{IJ} u_j \quad \text{if } i \in I, j \in J; \tag{5}$$

where $\{u_j, j \in \boldsymbol{I} - F\}$ are numbers satisfying

$$u_j > 0, \qquad \sum_{j \in J} u_j = 1. \tag{6}$$

Proof. Let

$$u_j = \sup_i u_{ij}.$$

From (1) and (3) we have

$$u_{ij} \leqq \sum_k u_{ik} u_j + u_{ij}(u_{jj} - u_j)$$

and so by (2)

$$u_{ij}(1 + u_j - u_{jj}) \leqq \left(\sum_k u_{ik}\right) u_j \leqq u_j.$$

Hence

$$u_j(1 + u_j - u_{jj}) \leqq u_j.$$

If $u_j = 0$ then $u_{jj} = 0$. If $u_j > 0$ then it follows from the last inequality that $u_j - u_{jj} \leqq 0$. Hence by the definition of u_j we have

$$u_{jj} = u_j. \tag{7}$$

Thus this equation holds in any case; using it and (3) again we obtain

$$u_{jj} \leqq \sum_k u_{jk} u_{jj} + u_{ji}(u_{ij} - u_{jj}) \leqq u_{jj} + u_{ji}(u_{ij} - u_{jj}).$$

It follows that

$$u_{ij} = u_{jj} \quad \text{if } u_{ji} > 0. \tag{8}$$

Now define F to be the set of j such that $u_{jj} = 0$. If $j \in F$, then $u_{ij} \leqq u_{jj} = 0$. Hence $u_{ij} = 0$ for every i, proving (4). Let $i \in \boldsymbol{I} - F$; we have

$$0 < u_{ii} \leqq \sum_j u_{ij} u_{ji} \leqq \left(\sum_j u_{ij}\right) u_i.$$

It follows by (7) that

$$\sum_j u_{ij} = 1 \quad \text{if } i \notin F. \tag{9}$$

Furthermore we have

$$1 = \sum_j u_{ij} \leqq \sum_j \sum_k u_{ik} u_{kj} = \sum_k u_{ik} \sum_j u_{kj} \leqq \sum_k u_{ik} = 1.$$

Thus equality must hold throughout, and we obtain

$$u_{ij} = \sum_k u_{ik} u_{kj} \quad \text{if } i \notin F,$$

strengthening (3). In particular

$$u_{ij} \geqq u_{ik} u_{kj};$$

consequently $u_{ik} > 0$ and $u_{kj} > 0$ imply that $u_{ij} > 0$. We may thus partition the set $\boldsymbol{I} - F$ into equivalence classes: two indices i and j belong to the same class if $u_{ij} > 0$. Thus relation is reflexive by (7) and symmetric by (8), since $u_{jj} > 0$ if $j \notin F$; and transitive as just shown. As a notational device, the class to which an index belongs will be denoted by the corresponding capital letter, e.g. $i \in I$, $j \in J$, unless the contrary is explicitly indicated. (The class I is not to be confused with $\boldsymbol{I}$.) Thus if $I \neq J$, then $u_{ij} = 0$. If $I = J$, then $u_{ji} > 0$ and we have from (7) and (8)

$$u_{ij} = u_{jj} = u_j.$$

These two cases are combined in the formula (5); (6) now follows from (9). ▯

The index set for the classes distinct from F will be denoted by $\boldsymbol{C}$ and an unspecified sum over the classes is over all $\boldsymbol{C}$.

Corollary. We have actually

$$u_{ij} = \sum_k u_{ik} u_{kj}, \tag{3'}$$

for every i and j. There exist numbers $\{\varrho_{iJ},\ i \in F,\ J \in \boldsymbol{C}\}$ satisfying

$$\varrho_{iJ} \geqq 0, \qquad \sum_J \varrho_{iJ} \leqq 1 \tag{10}$$

and such that

$$u_{ij} = \varrho_{iJ} u_j \quad \text{if } i \in F,\ j \in J. \tag{11}$$

Conversely, given any partition of $\boldsymbol{I}$ into disjoint sets $F, I, J, \ldots$ and given any $\{u_j, j \in \boldsymbol{I} - F\}$ and $\{\varrho_{iJ}, i \in F, J \in \boldsymbol{C}\}$, satisfying (6) and (10) respectively, there exists a matrix (u_{ij}) satisfying (1), (2), (3′), (4), (5) and (11).

Proof. By (4) the sum in (3) need only be extended to $k \notin F$. Now upon summing over all j and using (9) we see that equality must hold in (3), namely (3′) is true. We have from (3′), (4) and (5)

$$u_{ij} = \sum_k u_{ik}\, u_{kj} = \sum_{k \in J} u_{ik}\, u_{kj} = \Big(\sum_{k \in J} u_{ik}\Big)\, u_j .$$

If we define

$$\varrho_{iJ} = \sum_{k \in J} u_{ik}$$

then (10) is satisfied because of (1) and (2). To prove the converse we simply define the elements of the matrix (u_{ij}) by (4), (5) and (11). The verification of (1), (2) and (3′) is straightforward. ▯

We are now ready to prove that the continuity (hence also the measurability) of all p_{ij} is equivalent to the existence of their limits at zero; furthermore in this case each p_{ij} is decomposed into the product of this limit and a simpler function.

Theorem 3. Let (p_{ij}) be a transition matrix. Then all p_{ij} are continuous in $\boldsymbol{T}^0$ if and only if all the following limits exist:

$$\lim_{t \to 0} p_{ij}(t) = u_{ij} .$$

In this case the limit matrix (u_{ij}) satisfies the conditions (1), (2) and (3′) above and we have

$$(12) \qquad \sum_k u_{ik}\, p_{kj}(t) \leqq \sum_k p_{ik}(t)\, u_{kj} = p_{ij}(t) .$$

Furthermore, in the notation of Theorem 2, we have

$$(13) \qquad p_{ij}(t) \equiv 0 \quad \text{if } j \in F .$$

There exists a transition matrix (Π_{IJ}), $I, J \in \boldsymbol{C}$ satisfying the additional condition

$$(14) \qquad \lim_{t \to 0} \Pi_{IJ}(t) = \delta_{IJ}$$

such that

$$(15) \qquad p_{ij}(t) \equiv \Pi_{IJ}(t)\, u_j , \quad \text{if } i \in I,\ j \in J .$$

There exist functions Π_{iJ} on $\boldsymbol{T}^0$, $i \in F$, $J \in \boldsymbol{C}$ satisfying

$$(16) \qquad \begin{cases} \Pi_{iJ}(t) \geqq 0, & \sum_J \Pi_{iJ}(t) = 1 , \\ \sum_K \Pi_{iK}(s)\, \Pi_{KJ}(t) = \Pi_{iJ}(s+t) \end{cases}$$

and such that

$$(17) \qquad p_{ij}(t) \equiv \Pi_{iJ}(t)\, u_j \quad \text{if } i \in F,\ j \in J .$$

Conversely, given any partition of $\boldsymbol{I}$ into disjoint sets $F, I, J, \ldots$, and given any transition matrix (Π_{IJ}), $I, J \in \boldsymbol{C}$ satisfying (14), any continuous functions $\{\Pi_{iJ}, i \in F, J \in \boldsymbol{C}\}$ satisfying (16) and any $\{u_j, j \in \boldsymbol{I}-F\}$ satisfying (6), there exists a measurable transition matrix satisfying (13), (15) and (17).

Proof. Suppose that all p_{ij} are continuous. Let $\{t_n\}$, $\{t'_n\}$ be two sequences tending to zero such that $\lim\limits_{n\to\infty} p_{ij}(t_n) = u_{ij}$, $\lim\limits_{n\to\infty} p_{ij}(t'_n) = u'_{ij}$ for all i, j. Such sequences exist be the Bolzano-Weierstrass theorem and the diagonal procedure. We have by Fatou's lemma,

$$1 = \lim_{n\to\infty} \sum_j p_{ij}(t_n) \geqq \sum_j u_{ij}.$$

Furthermore, by the continuity of p_{ij} and dominated convergence we have

$$p_{ij}(t) = \lim_{n\to\infty} p_{ij}(t+t_n) = \lim_{n-\infty} \sum_k p_{ik}(t)\, p_{kj}(t_n) = \sum_k p_{ik}(t)\, u_{kj}; \tag{18}$$

and consequently

$$1 = \sum_j p_{ij}(t) = \sum_k p_{ik}(t) \sum_j u_{kj}.$$

It follows that if $\sum\limits_j u_{kj} < 1$ for a certain k, then $p_{ik}(t) = 0$ for this k, every i and every t; in particular $u'_{ik} = 0$. Next we have by Fatou's lemma

$$p_{ij}(t) = \lim_{n\to\infty} \sum_k p_{ik}(t'_n)\, p_{kj}(t) \geqq \sum_k u'_{ik}\, p_{kj}(t). \tag{19}$$

Letting $t \to 0$ along the sequence $\{t_n\}$ we obtain

$$u_{ij} \geqq \sum_k u'_{ik}\, u_{kj}. \tag{20}$$

On the other hand, letting $t \to 0$ along $\{t'_n\}$ in (18) we obtain

$$u'_{ij} \geqq \sum_k u'_{ik}\, u_{kj}; \tag{21}$$

and consequently

$$\sum_j u'_{ij} \geqq \sum_k u'_{ik} \sum_j u_{kj} = \sum_k u'_{ik}$$

since $u'_{ik} = 0$ whenever $\sum\limits_j u_{kj} < 1$, as shown above. It follows that equality must hold in (21), and upon comparison with (20) we obtain $u_{ij} \geqq u'_{ij}$ and so $u_{ij} = u'_{ij}$ by symmetry. Thus any two limiting matrices of $(p_{ij}(t))$ as $t \to 0$ are identical and there is a unique limit matrix (u_{ij}). The relation (21) with equality sign becomes (3′) and (18) and (19) become (12). We have therefore proved that continuity of all p_{ij} implies the existence of a unique limit matrix (u_{ij}) at zero satisfying the stated conditions. Suppose on the other hand that there is a unique limit matrix (u_{ij}). Then reading (18) without the first member we see that every p_{ij} has a right-hand limit at every t. Such a function has only a denumerable set of discontinuities (see e.g. SAKS [1; p. 261]). Hence all p_{ij} are measurable and therefore continuous by Theorem 1.

We proceed to prove the other assertions. If $j\in F$, then $u_{ij}=0$ for every i by (4) and it follows from (12) that $p_{ij}(t)\equiv 0$. Let $i\in I$, $j\in J$. Since $p_{ik}(t)\,u_{kj}=0$ if $k\notin J$, and $u_{kj}=u_j$ if $k\in J$ by (5), we have from (12)

$$p_{ij}(t)=\Bigl(\sum_{k\in J}p_{ik}(t)\Bigr)u_j.$$

Thus $p_{ij}(t)\,u_j^{-1}$ depends only on i and J. Furthermore, upon summing (19) and writing u_{ik} for u'_{ik} we have

$$1\geqq\sum_k u_{ik}\sum_j p_{kj}(t)\geqq\sum_k u_{ik}=1,$$

the last equation following from (9). Hence equality must hold in (19) and using (5) we obtain

$$p_{ij}(t)=\sum_{k\in I}u_k\,p_{kj}(t).$$

Thus $p_{ij}(t)\,u_j^{-1}$ depends only on I and j. Combining this with the preceding remark we conclude that $p_{ij}(t)\,u_j^{-1}$ depends only on I and J so that we can define Π_{IJ} by (15). Clearly $\Pi_{IJ}(t)\geqq 0$;

$$1=\sum_{j\notin F}p_{ij}(t)=\sum_J\sum_{j\in J}\Pi_{IJ}(t)\,u_j=\sum_J\Pi_{IJ}(t);$$

and

$$\begin{aligned}\Pi_{IJ}(s+t)=p_{ij}(s+t)\,u_j^{-1}&=\sum_{k\notin F}p_{ik}(s)\,p_{kj}(t)\,u_j^{-1}=\sum_{k\notin F}\Pi_{IK}(s)\,u_k\,\Pi_{KJ}(t)\\&=\sum_K\Pi_{IK}(s)\,\Pi_{KJ}(t)\sum_{k\in K}u_k=\sum_K\Pi_{IK}(s)\,\Pi_{KJ}(t).\end{aligned}$$

Finally we have by (5),

$$\lim_{t\to 0}\Pi_{IJ}(t)=\lim_{t\to 0}p_{ij}(t)\,u_j^{-1}=u_{ij}\,u_j^{-1}=\delta_{IJ}.$$

We have thus verified all the properties asserted for (Π_{IJ}).

Next, we have from (18), since $p_{ik}(t)=0$ if $k\in F$ and $u_{kj}=0$ if $k\notin J$,

$$p_{ij}(t)=\Bigl(\sum_{k\in J}p_{ik}(t)\Bigr)u_j.$$

We define

$$\Pi_{iJ}(t)=\sum_{k\in J}p_{ik}(t)\quad\text{if } i\in F,\ J\in\mathbf{C}$$

so that (17) is true. Then we have $\Pi_{iJ}(t) \geqq 0$;

$$\sum_J \Pi_{iJ}(t) = \sum_J \sum_{j\in J} p_{ij}(t) = \sum_{j\notin F} p_{ij}(t) = 1;$$

$$\Pi_{iJ}(s+t) = p_{ij}(s+t)\,u_j^{-1} = \sum_k p_{ik}(s)\,p_{kj}(t)\,u_j^{-1}$$

$$= \sum_K \sum_{k\in K} \Pi_{iK}(s)\,u_k\,\Pi_{KJ}(t) = \sum_K \Pi_{iK}(s)\,\Pi_{KJ}(t).$$

Thus all the relations in (16) are satisfied.

The proof of the converse is straightforward and will be left to the reader. ▯

Corollary. If (p_{ij}) is measurable and $i\notin F$, then the series in (B) converges uniformly in every finite t-interval.

Proof. Extend p_{ij} from $\boldsymbol{T}^0$ to $\boldsymbol{T}$ by defining $p_{ij}(0) = u_{ij}$. The corollary then follows from (9) and DINI's theorem on uniform convergence (see e.g. TITCHMARSH [1; p. 13]). ▯

Theorem 4. Let (p_{ij}) be a measurable transition matrix. We have $p_{jj}(t) \equiv 0$ or $p_{jj}(t) > 0$ for all t according as $j\in F$ or $j\notin F$. If $p_{ij}(t_0) > 0$ then $p_{ij}(t) > 0$ for all $t \geqq t_0$.

Proof. If $j\in F$, then $p_{jj}(t) \equiv 0$ by (13). If $j\notin F$, then by (5) and (15), $\lim_{t\to 0} p_{jj}(t) = u_j > 0$. Since

$$p_{jj}(t) \geqq \left[p_{jj}\left(\frac{t}{n}\right)\right]^n \tag{22}$$

for every n by (A) and (C), it follows that $p_{jj}(t) > 0$ for all t. Next, $p_{ij}(t_0) > 0$ implies that $j\notin F$; hence if $t > t_0$ we have by (A) and (C) and what has just been proved,

$$p_{ij}(t) \geqq p_{ij}(t_0)\,p_{jj}(t - t_0) > 0. \quad ▯$$

The second assertion of Theorem 4 will now be sharpened into a rather deep result.

Theorem 5. Let (p_{ij}) be a measurable transition matrix. Then each p_{ij} is either identically zero or never zero in $\boldsymbol{T}^0$.

Proof. We may take $\boldsymbol{I}$ to be the set of positive integers. Suppose first that the theorem is false for p_{il} where $i\notin F$. Then by Theorem 4 there exists an $t_0 > 0$ such that

$$p_{il}(t) = 0 \quad \text{if } 0 < t \leqq t_0 \quad \text{and} \quad p_{il}(2t_0) = c > 0.$$

By the Corollary to Theorem 3, there exists an N such that

$$\sum_{j>N} p_{ij}(t) < \frac{c}{4} \quad \text{if } 0 < t \leqq 2t_0. \tag{23}$$

Let $s = t_0/2N$ and define for $m \geqq 1$:

$$A_m = \{k: p_{ik}(m\,s) > 0\}.$$

By Theorem 4 we have $A_m \subset A_{m+1}$. Let $B_1 = A_1$, $B_m = A_m - A_{m-1}$ for $m \geqq 2$. If $k \notin A_m$, then

$$0 = p_{ik}(m\,s) = \sum_j p_{ij}\big((m-1)\,s\big)\,p_{jk}(s) = \sum_{j \in A_{m-1}} p_{ij}\big((m-1)\,s\big)\,p_{jk}(s)$$

and consequently

$$p_{jk}(s) = 0 \quad \text{if } j \in A_{m-1},\ k \notin A_m. \tag{24}$$

Suppose if possible that $A_m = A_{m-1}$ for a certain m, $1 < m \leqq 2N$. Then it follows from (24) that

$$p_{ik}\big((m+1)\,s\big) = \sum_{j \in A_m} p_{ij}(m\,s)\,p_{jk}(s) = 0 \quad \text{if } k \notin A_m.$$

Thus $A_{m+1} = A_m$ from which we obtain, by repeating the argument, $A_{m'} = A_m$ for all $m' > m$. This is impossible since $l \notin A_{2N}$ but $l \in A_{4N}$ by the hypothesis on p_{il}. We have therefore proved that the sets B_m, $1 \leqq m \leqq 2N$, are all nonempty and disjoint.

Let $1 \leqq m \leqq 2N$; if $k \notin A_m$ then we have by (24), for every $n \geqq 1$,

$$p_{ik}\big((n+1)\,s\big) = \Big(\sum_{j \notin A_m} + \sum_{j \in B_m} + \sum_{j \in A_{m-1}}\Big) p_{ij}(n\,s)\,p_{jk}(s)$$

where the third sum vanishes on account of (24). Hence, summing over all $k \notin A_m$, we obtain

$$\sum_{k \notin A_m} p_{ik}\big((n+1)\,s\big) \leqq \sum_{j \notin A_m} p_{ij}(n\,s) + \sum_{j \in B_m} p_{ij}(n\,s).$$

Summing over n from 0 to $4N-1$, we obtain

$$\sum_{k \notin A_m} p_{ik}(4N\,s) \leqq \sum_{n=1}^{4N} \sum_{j \in B_m} p_{ij}(n\,s). \tag{25}$$

Since $l \notin A_{2N}$, the left member is at least equal to $p_{il}(4N\,s) = c$. Thus we have

$$c \leqq \sum_{n=1}^{4N} \sum_{j \in B_m} p_{ij}(n\,s), \qquad 1 \leqq m \leqq 2N.$$

Since the sets $B_1, \ldots, B_{2N}$ are nonempty and disjoint, there exist at least N of them, whose union we denote by B, which are disjoint from the set $(1, \ldots, N)$. We have

$$N\,c \leqq \sum_{n=1}^{4N} \sum_{j \in B} p_{ij}(n\,s). \tag{26}$$

On the other hand, we have from (23),

$$\sum_{j \in B} p_{ij}(n\,s) \leqq \sum_{j > N} p_{ij}(n\,s) < c/4$$

and consequently the right member of (26) is strictly less than $4N(c/4)=Nc$. This contradiction proves the assertion of the theorem in case $i\notin F$.

Now let i be arbitrary and suppose that $p_{ij}(t_1)>0$ for a certain $t_1>0$. Then by (13) there is at least one k not in F such that $p_{ik}(t_1/2)>0$ and $p_{kj}(t_1/2)>0$. It follows from what has just been proved that $p_{ij}(3t_1/4)>0$. Repeating this argument and using Theorem 4 we conclude that $p_{ij}(t)>0$ for all $t>0$, as was to be shown. ▯

Another completely different and probabilistic proof of Theorem 5 will be given in Theorem 5.2.

Notes. Theorems 1, 2 and 3 are due to DOOB [1], given here with slight improvements. Similar results in the case where $\boldsymbol{I}$ is finite were previously given by DOEBLIN [2]. The proof of Theorem 1 given here follows the general semigroup treatment (see HILLE-PHILLIPS [1, p. 305]) with a simplification communicated by JURKAT, who also observed that (3′) follows from (1) to (3). The proof of Theorem 2 is that of CHACON [1]. Both are simpler than DOOB's original versions. Theorem 5 was first proved by D. G. AUSTIN and simplified by the author, using probabilistic methods; see Theorem 5.2 below. The algebraic proof given here was communicated by D. ORNSTEIN. The result remains true even if condition (B) for (p_{ij}) is dropped; see CHUNG [2].

The term "transition matrix" as defined in this section should not cause confusion with that in § I.2, where it denotes the stochastic matrix of the one-step transition probabilities of a discrete parameter M.C. In the latter sense the term may be re-named "*one-step transition matrix*" to avoid any conflict. Alternatively, the more logical but academic term "*transition matrix function*" may be used for the present purpose, as in DOOB [4].

§ 2. Standard transition matrix

According to Theorem 1.3, the analytic study of (p_{ij}) is reduced to that of (Π_{IJ}) if the set of indices F is ignored. In fact, the curtailed matrix (p_{ij}), $i,j\in\boldsymbol{I}-F$, is a transition matrix and its elements differ from those of (Π_{IJ}) only in certain constant factors depending on the second index. From the standpoint of probability, it will be seen (in §4) that the set F plays a nuisance role and can indeed be ignored. Moreover, the reduction from (p_{ij}) to (Π_{IJ}) is also justified on probabilistic grounds (see Theorem 4.3). The distinctive feature of the transition matrix (Π_{IJ}) is the property (1.14) which will now be formulated as a definition.

A transition matrix (p_{ij}) is called *standard* iff

$$\text{(E)}\qquad \lim_{t\to 0} p_{ij}(t)=\delta_{ij}$$

for every i and j. This is equivalent to $\lim_{t\to 0} p_{ii}(t)=1$ for every i. For a standard transition matrix it is natural to extend the definition of each p_{ij} to $\boldsymbol{T}$ by setting

$$p_{ij}(0)=\delta_{ij}. \tag{1}$$

The following result is a simple consequence of Theorem 1.3.

Theorem 1. The transition matrix (p_{ij}) is standard if and only if

(a) it is measurable;

(b) in the notation of §1, the class F is empty and each of the remaining classes contains only one element.

In particular, the continuity of the elements of a standard transition matrix follows from Theorem 1.3, but a sharper result can be proved in a simpler way.

Theorem 2. Let (p_{ij}) be a standard transition matrix. Then for every i, j, $t\geqq 0$ and $h>0$ we have

$$|p_{ij}(t+h)-p_{ij}(t)|\leqq 1-p_{ii}(h). \tag{2}$$

In other words, the modulus of continuity of p_{ij} does not exceed that of p_{ii} at zero.

Proof. We have by (1.C),

$$\begin{aligned} p_{ij}(t+h)-p_{ij}(t) &= \sum_k p_{ik}(h)p_{kj}(t)-p_{ij}(t) \\ &= [p_{ii}(h)-1]p_{ij}(t)+\sum_{k\neq i} p_{ik}(h)p_{kj}(t). \end{aligned}$$

Neither term in the last member exceeds $1-p_{ii}(h)=\sum_{k\neq i} p_{ik}(h)$ in absolute value by (1.A) and (1.B) and they are of opposite signs; hence (2) follows. ▯

Corollary. Let (p_{ij}) be a measurable transition matrix. For every $i\notin F$ and every j, p_{ij} is uniformly continuous in $\boldsymbol{T}$.

It turns out that the continuity of p_{ij} follows already from assumptions (1.A), (1.C) and (2.E) without the intervention of (1.B), the finiteness of all p_{ij} being assumed. We prove the following more general result which will also be used on several other occasions; note that the dual statements correspond to the transposition of a matrix.

Theorem 3. Let (g_{ij}), $i,j\in\boldsymbol{I}$, be a matrix of finite, nonnegative functions on $\boldsymbol{T}$ satisfying the condition that for every i:

$$\lim_{t\downarrow 0} g_{ii}(t)=1. \tag{3}$$

Let $\{f_j, j \in \boldsymbol{I}\}$ be finite, nonnegative functions on $\boldsymbol{T}^0$ satisfying the following equations for s, $t \in \boldsymbol{T}^0$:

$$f_j(s+t) = \sum_i f_i(s) g_{ij}(t), \quad j \in \boldsymbol{I}; \tag{4}$$

$$\left[\text{or } f_i(s+t) = \sum_j g_{ij}(s) f_j(t), \quad i \in \boldsymbol{I}\right]. \tag{5}$$

Then each f_j may be extended to $\boldsymbol{T}$ to be continuous there.

Proof. We prove (4) since (5) is entirely similar. For a fixed j and $t \geqq 0$ suppose that $t_n \to 0$, $t'_n \to 0$, $0 < t'_n < t_n$, $f_j(t+t_n) \to a_j$, $f_j(t+t'_n) \to a'_j$. Then it follows from (4) that

$$f_j(t+t_n) = \sum_i f_i(t+t'_n) g_{ij}(t_n - t'_n) \geqq f_j(t+t'_n) g_{jj}(t_n - t'_n)$$

so that, upon letting $n \to \infty$ and using (3), we have $a_j \geqq a'_j$. Hence by symmetry $a_j = a'_j$ and we have proved that $f_j(t+0)$ exists; similarly $f_j(t-0)$ exists if $t > 0$. The inequality above shows also that these limits are finite everywhere. Such a function f_j has at most a countable set D_j of discontinuities (see e.g. Saks [1; p. 261]); put $D = \bigcup_{j \in \boldsymbol{I}} D_j$. For each $\varepsilon > 0$ we have also $f_j(t+\varepsilon) \geqq f_j(t) g_{jj}(\varepsilon)$ from which and a similar argument it follows that for every j:

$$f_j(t+0) \geqq f_j(t) \geqq f_j(t-0). \tag{6}$$

Given any $t > 0$, there exists an s such that $0 < s < t$ and $s \notin D$. If $0 < \varepsilon < s$ then

$$f_j(t-\varepsilon) = \sum_i f_i(s-\varepsilon) g_{ij}(t-s),$$

and consequently by Fatou's lemma

$$f_j(t-0) \geqq \sum_i f_i(s) g_{ij}(t-s) = f_j(t).$$

Together with (6) we see that each f_j is left continuous. If D were not empty, then this and (6) would imply that for any t_0 in D, there is some j such that $f_j(t_0+0) > f_j(t_0)$. It follows that for all s so small that $g_{jj}(s) > 0$, we should have

$$\begin{aligned} f_j(t_0+s+0) &= \overline{\lim_{\delta \downarrow 0}}\, f_j(t_0+s+\delta) \\ &\geqq \overline{\lim_{\delta \downarrow 0}}\, f_j(t_0+\delta) g_{jj}(s) + \underline{\lim}_{\delta \downarrow 0} \sum_{i \neq j} f_i(t_0+\delta) g_{ij}(s) \\ &\geqq f_j(t_0+0) g_{jj}(s) + \sum_{i \neq j} f_i(t_0) g_{ij}(s) \\ &> \sum_i f_i(t_0) g_{ij}(s) = f_j(t_0+s). \end{aligned}$$

Hence all such s would belong to D which is impossible since D is countable. Thus D is empty and we have proved that each f_j is continuous in $\boldsymbol{T}^0$ and has a finite limit at 0. The theorem follows. ▯

Corollary. Let (g_{ij}) be finite, nonnegative functions on $\boldsymbol{T}^0$ satisfying the condition (3) and the following equations:

$$g_{ij}(s+t)=\sum_k g_{ik}(s)g_{kj}(t), \qquad s, t\in\boldsymbol{T}^0.$$

Then each g_{ij} is continuous in $\boldsymbol{T}$ by extension.

Remark. Uniform continuity in $\boldsymbol{T}$ is ruled out by the trivial counterexample $g_{ij}(t)=\delta_{ij}\,e^{q_i t}$ where $q_i>0$.

We proceed to establish the existence of the derivatives at zero of the elements of a standard transition matrix (p_{ij}) in the next two theorems. Naturally we consider only the right-hand derivative at zero which will be denoted simply by $p'_{ij}(0)$.

Theorem 4. For every i,

$$-p'_{ii}(0)=\lim_{t\to 0}\frac{1-p_{ii}(t)}{t}$$

exists but may be infinite.

Proof. We know from Theorem 1.4 that $p_{ii}(t)>0$ for all t; let

$$-\log p_{ii}(t)=\varphi(t);$$

then φ is finite-valued. The inequality

$$p_{ii}(s+t)\geqq p_{ii}(s)p_{ii}(t) \tag{7}$$

which follows from (1.A) and (1.C), becomes the *subadditive* property of φ:

$$\varphi(s+t)\leqq\varphi(s)+\varphi(t).$$

Put

$$q_i=\sup_{t>0}\frac{\varphi(t)}{t}\leqq\infty.$$

If $q_i<\infty$, there exists a t_0 such that $\varphi(t_0)/t_0>q_i-\varepsilon$. For each t write $t_0=n\,t+\delta$ where $0\leqq\delta<t$; we have

$$q_i-\varepsilon\leqq\frac{\varphi(t_0)}{t_0}\leqq\frac{n\varphi(t)+\varphi(\delta)}{t_0}=\frac{nt}{t_0}\,\frac{\varphi(t)}{t}+\frac{\varphi(\delta)}{t_0}.$$

As $t\to 0$, $nt/t_0\to 1$ and $\varphi(\delta)\to 0$ by (E); hence

$$q_i-\varepsilon\leqq\varliminf_{t\to 0}\frac{\varphi(t)}{t}\leqq\varlimsup_{t\to 0}\frac{\varphi(t)}{t}\leqq q_i.$$

Since ε is arbitrary we have

$$\lim_{t\to 0} \frac{\varphi(t)}{t} = q_i.$$

If $q_i = \infty$ we need only replace the $q_i - \varepsilon$ above by M where M is any given large number, and the same conclusion holds. We have by (1),

$$\lim_{t\to 0} \frac{\varphi(t)}{t} = \lim_{t\to 0} \frac{-\log[1-(1-p_{ii}(t))]}{t} = \lim_{t\to 0} \frac{1-p_{ii}(t)}{t}.$$

Hence the limit exists. ▯

Remark. Observe that the above proof uses only the finiteness of p_{ii} and (7), the latter being a consequence of (1.A) and (1.C) without the use of (1.B).

Theorem 5. For every i and j, $i \neq j$,

$$p'_{ij}(0) = \lim_{t\to 0} \frac{p_{ij}(t)}{t} \tag{8}$$

exists and is finite.

Proof. In this proof we use an essentially probabilistic argument based on the properties of discrete parameter Markov chains discussed in Part I. For each fixed $h>0$ the matrix $(p_{ij}(h))$ is the transition matrix of the discrete parameter M.C. $\mathfrak{C}_h = \{x_{nh}, n \geqq 0\}$ which is the h-skeleton of a continuous parameter M.C. $\{x_t, t \geqq 0\}$ with the given $(p_{ij}(\cdot))$ as transition matrix (see §§ 4, 10 below), but it is unnecessary for the present proof to introduce the latter. Using the notation of § I.9, but with h in parenthesis, we define ${}_jp_{ii}^{(0)}(h) = 1$ and

$$\begin{aligned} {}_jp_{ii}^{(n)}(h) &= \boldsymbol{P}\{x_{nh}(\omega)=i;\ x_{\nu h}(\omega) \neq j,\ 1 \leqq \nu < n \mid x_0(\omega)=i\}, \\ f_{ij}^{(n)}(h) &= \boldsymbol{P}\{x_{nh}(\omega)=j;\ x_{\nu h}(\omega) \neq j,\ 1 \leqq \nu < n \mid x_0(\omega)=i\}. \end{aligned}$$

We have then

$$p_{ij}(nh) \geqq \sum_{\nu=0}^{n-1} {}_jp_{ii}^{(\nu)}(h)\, p_{ij}(h)\, p_{jj}((n-\nu-1)h) \tag{9}$$

since each term on the right side represents a possible way of going from i to j in n steps in $\mathfrak{C}_h$ and different terms correspond to mutually exclusive events. Furthermore, we have by an application of (I.9.1),

$$p_{ii}(\nu h) = {}_jp_{ii}^{(\nu)}(h) + \sum_{m=1}^{\nu-1} f_{ij}^{(m)}(h)\, p_{ji}((\nu-m)h). \tag{10}$$

Since

$$\sum_{m=1}^{\nu-1} f_{ij}^{(m)}(h) \leqq 1, \tag{11}$$

we have

$$(12)\qquad {}_jp_{ii}^{(\nu)}(h) \geqq p_{ii}(\nu h) - \max_{1\leqq m\leqq \nu} p_{ji}\big((\nu-m)h\big).$$

It follows from property (E) of a standard transition matrix that given any $\varepsilon > 0$, there exists a t_0 such that (for the given i and j)

$$\max_{0\leqq t\leqq t_0} p_{ji}(t) < \varepsilon, \qquad \min_{0\leqq t\leqq t_0} p_{ii}(t) > 1-\varepsilon, \qquad \min_{0\leqq t\leqq t_0} p_{jj}(t) > 1-\varepsilon.$$

Hence if $n h < t_0$, then ${}_jp_{ii}^{(\nu)}(h) > 1-2\varepsilon$ by (12). Using this in (9) we obtain

$$p_{ij}(n h) > (1-2\varepsilon)\sum_{\nu=0}^{n-1} p_{ij}(h)(1-\varepsilon) \geqq (1-3\varepsilon)\, n\, p_{ij}(h);$$

or

$$(13)\qquad \frac{p_{ij}(nh)}{nh} > (1-3\varepsilon)\frac{p_{ij}(h)}{h}, \quad \text{if } n h < t_0.$$

Put

$$q_{ij} = \overline{\lim_{t\to 0}}\, \frac{p_{ij}(t)}{t}.$$

Letting $h\to 0$ and $n h\to t$ where $0 < t < t_0$ in (13), we obtain by the continuity of p_{ij}:

$$\frac{p_{ij}(t)}{t} \geqq (1-3\varepsilon)\, q_{ij}.$$

It follows that $q_{ij} < \infty$ and

$$\underline{\lim_{t\to 0}}\, \frac{p_{ij}(t)}{t} \geqq q_{ij}$$

since ε is arbitrary. This proves the theorem. ▯

Remark. The introduction of the M.C. $\mathfrak{C}_h$ is for the sake of motivation of the formulas. The quantities ${}_jp_{ii}^{(n)}(h)$ and $f_{ij}^{(n)}(h)$ can of course be defined algebraically as follows:

$$ {}_jp_{ii}^{(n)}(h) = \sum_{k_1\neq j}\ \sum_{k_2\neq j}\cdots\sum_{k_{n-1}\neq j} p_{ik_1}(h)\, p_{k_1k_2}(h)\, \dots\, p_{k_{n-1}i}(h);$$

$$ f_{ij}^{(n)}(h) = \sum_{k_1\neq j}\ \sum_{k_2\neq j}\cdots\sum_{k_{n-1}\neq j} p_{ik_1}(h)\, p_{k_1k_2}(h)\, \dots\, p_{k_{n-1}j}(h).$$

The formulas (9) and (10) can be verified using only properties (1.A) and (1.C) of the transition matrix (p_{ij}). In order to verify (11) for sufficiently small h (which is all that is necessary), without using (1.B), we note the relation

$$p_{ij}(\nu h) = \sum_{m=1}^{\nu} f_{ij}^{(m)}(h)\, p_{jj}\big((\nu-m)h\big)$$

which follows from (1.C). Chose t_0 in the above proof to satisfy also

$$\max_{0\leqq t\leqq t_0} p_{ij}(t) < \varepsilon;$$

then since $\nu h < t_0$ we have

$$\sum_{m=1}^{\nu} f_{ij}^{(m)}(h)(1-\varepsilon) < \varepsilon .$$

Thus (11) is true if $\varepsilon < \frac{1}{2}$. We have therefore shown that Theorem 5 is true under the assumptions (1.A), (1.C) and (2.E), the finiteness of all p_{ij} being assumed.

Theorem 6. For every i,

$$\sum_{j \neq i} p'_{ij}(0) \leqq -p'_{ii}(0) \leqq \infty . \tag{14}$$

Proof. Write

$$\frac{1-p_{ii}(t)}{t} - \sum_{j \neq i} \frac{p_{ij}(t)}{t} = \frac{1-\sum\limits_{j} p_{ij}(t)}{t} = 0 ,$$

and let $t \to 0$. Each term on the left side has a limit by Theorems 4 and 5. The inequality follows from FATOU's lemma. (For examples where strict inequality holds see Examples 20.2 to 20.4.) ▯

Notes. The term "standard transition matrix" is new; most authors are concerned only with such a matrix so that no name has been proposed for it. The uniform continuity in Theorem 2 was stated by LÉVY [1], who used it to prove Theorem 10.1 below. Theorem 3 is new. Theorem 4 is known in the theory of subadditive functions; e.g. HILLE-PHILLIPS [1; p. 250]. Theorem 5 is due to KOLMOGOROV [5]; previously DOOB [1] proved (8) under the hypothesis that $-p'_{ii}(0) < \infty$. No "purely analytic" proof of this result is known. To W. JURKAT is due the remarkable observation that assumption (1.B) can be dispensed with in several theorems of this and the next section; see JURKAT [1]. For generalization of Theorems 4 and 5 to "discontinuous Markov processes" in a general state space see D. G. KENDALL [1].

§ 3. Differentiability

Let (p_{ij}) be a standard transition matrix. The derivatives at zero of the p_{ij} established in the preceding section are of basic importance in the study of the associated Markov chain. The following notation will be used throughout the rest of this monograph:

$$q_i = -p'_{ii}(0), \quad q_{ij} = p'_{ij}(0), \quad i \neq j . \tag{1}$$

Occasionally the notation $q_{ii} = -q_i$ will also be used; the matrix

$$(q_{ij}) = (p'_{ij}(0))$$

will then be called the *Q-matrix* of the matrix (p_{ij}).

In the next two theorems it will be proved that if either $q_i < \infty$ or $q_j < \infty$ then p_{ij} has a continuous derivative in $\boldsymbol{T}$. A more general result will be given in § 12.

Theorem 1. If $q_i < \infty$, then for every j, p_{ij} has a continuous derivative in $\boldsymbol{T}$. Furthermore, we have

$$|p_{ij}(t+h) - p_{ij}(t)| \leqq q_i h, \qquad t \geqq 0,\ h \geqq 0; \tag{2}$$

$$\sum_j p'_{ij}(t) = 0, \qquad t > 0; \tag{3}$$

$$\sum_j |p'_{ij}(t)| \leqq 2q_i; \qquad t \geqq 0; \tag{4}$$

$$p'_{ij}(s+t) = \sum_k p'_{ik}(s) p_{kj}(t), \qquad s > 0,\ t \geqq 0. \tag{5}$$

Proof. It follows from the definition of q_i and (1.22) that

$$p_{ii}(t) \geqq e^{-q_i t} \geqq 1 - q_i t \tag{6}$$

and so (2) follows from (2.2). The inequality (2) means that the function p_{ij} satisfies the Lipschitz condition; it is therefore absolutely continuous.

If $q_i = 0$ then (6) implies that $p_{ii}(t) \equiv 1$ so that $p_{ij}(t) \equiv \delta_{ij}$ and all the other assertions of the theorem are trivial. Hence we may suppose that $q_i > 0$. Using (1.C) and (6) we have

$$p_{ij}(t+h) - p_{ij}(t) \geqq [p_{ii}(h) - 1] p_{ij}(t) \geqq -q_i h\, p_{ij}(t)$$

and consequently

$$\underline{D}^+ [e^{q_i t} p_{ij}(t)] = e^{q_i t} [\underline{D}^+ p_{ij}(t) + q_i p_{ij}(t)] \geqq 0 \tag{7}$$

where $\underline{D}^+$ denotes the right-hand lower derivate. This and the continuity of p_{ij} imply, by DINI's theorem on derivates (see e.g. SAKS [1; p. 204]), that $e^{q_i t} p_{ij}(t)$ is nondecreasing in t and hence p_{ij} has a finite derivative Dp_{ij} a.e. (The latter assertion follows of course already from the Lipschitzian character of p_{ij}.) Put

$$r_{ij}(t) = q_i^{-1} D p_{ij}(t) + p_{ij}(t) \geqq 0 \tag{8}$$

so that r_{ij} is defined wherever Dp_{ij} is. Let us rewrite (1.B) and (1.C) as follows:

$$\sum_j e^{q_i s} p_{ij}(s) = e^{q_i s};$$

$$e^{q_i (s+t)} p_{ij}(s+t) = e^{q_i t} \sum_k e^{q_i s} p_{ik}(s) p_{kj}(t).$$

Now a theorem of Fubini asserts: If $f = \sum_n f_n$ is a convergent series of nondecreasing functions, then the relation $f' = \sum_n f_n'$ holds a.e. (see e.g. SAKS [1; p. 117]). Applying this result to the two equations above, keeping t fixed, we obtain

$$\sum_j r_{ij}(s) = 1, \tag{9}$$

$$r_{ij}(s+t) = \sum_k r_{ik}(s) p_{kj}(t). \tag{10}$$

These equations are valid for each t and a.a. s, with the exceptional set depending on t. Since the members are measurable functions of the pair (s, t), the equations are also valid, by FUBINI's theorem on product measure, if $s \notin Z$ and $t \notin Z_s$ where $\mu(Z) = \mu(Z_s) = 0$. Fix an $s_0 \notin Z$; then (9) implies that the right member of (10) is continuous in t and so $r_{ij}(s_0+t)$ as a function of t coincides with a continuous function a.e. in $\boldsymbol{T}^0$. Hence one version of the a.e. derivative $Dp_{ij}(s_0+t)$ is a continuous function of t and therefore $p_{ij}(s_0+t)$, being absolutely continuous and so the integral of its a.e. derivative, possesses in fact a continuous derivative everywhere in $\boldsymbol{T}^0$. This being true for every $s_0 \notin Z$ we conclude that p_{ij} has a continuous derivative in $\boldsymbol{T}^0$, proving the main assertion of the theorem. It follows that r_{ij} is defined and continuous in $\boldsymbol{T}^0$, for every j. Returning to (9) and (10), we now show that these equations are valid for all $s, t \in \boldsymbol{T}^0$. For each $s \notin Z$, both members of (10) are continuous in t; hence (10) holds for all $t \in \boldsymbol{T}^0$. Next, let $s > 0$ be arbitrary and choose $0 < s' < s$, $s' \notin Z$. We have

$$\begin{aligned} r_{ij}(s+t) &= r_{ij}(s' + (s - s' + t)) \\ &= \sum_k r_{ik}(s') \sum_l p_{kl}(s-s') p_{lj}(t) = \sum_l r_{il}(s) p_{lj}(t). \end{aligned}$$

Thus (10) holds for all $s > 0$ and $t \geqq 0$. Summing it over all j, we obtain

$$\sum_j r_{ij}(s+t) = \sum_k r_{ik}(s).$$

Hence (9) holds for all $s > 0$.

Substituting (8) into (9) and (10) we obtain (3) and (5). Furthermore we have from (8),

$$|p_{ij}'(t)| \leqq q_i [r_{ij}(t) + p_{ij}(t)]$$

and consequently (4) follows from (9). Finally, the existence of $\lim_{t \to 0} p_{ij}'(t)$ is equivalent to that of $\lim_{t \to 0} r_{ij}(t)$ which follows from (10) and Theorem 2.3. We have then $\lim_{t \to 0} p_{ij}'(t) = p_{ij}'(0) = q_{ij}$ by the mean value theorem of differential calculus. Hence p_{ij}' is continuous in $\boldsymbol{T}$. ▯

Corollary. The sum of the series $\sum_j |p'_{ij}(t)|$ is nonincreasing in t; the series converges uniformly in $t \geqq \delta > 0$; and we have

$$\lim_{h\to 0} \sum_j |p'_{ij}(t+h) - p'_{ij}(t)| = 0;$$

$$\lim_{h\to 0} \sum_j \left| \frac{p_{ij}(t+h) - p_{ij}(t)}{h} - p'_{ij}(t) \right| = 0$$

uniformly in $t \geqq \delta > 0$. If (3) holds also at $t = 0$, namely if $q_i = \sum_{j\neq i} q_{ij}$, then the uniformity holds in $t \geqq 0$ in all three cases.

The proof of the corollary is similar to that of Theorem 1.1.

Theorem 2. If $q_j < \infty$, then for every i, p_{ij} has a continuous derivative in $\boldsymbol{T}$. Furthermore we have

$$p'_{ij}(t+s) = \sum_k p_{ik}(t)\, p'_{kj}(s), \qquad t \geqq 0,\ s > 0. \tag{11}$$

Proof. We have, using (6),

$$p_{ij}(t+h) - p_{ij}(t) \geqq p_{ij}(t)\,[p_{jj}(h) - 1] \geqq -p_{ij}(t)\, q_j\, h$$

and consequently

$$D[p_{ij}(t)\, e^{q_j t}] = [D p_{ij}(t) + p_{ij}(t)\, q_j]\, e^{q_j t} \geqq 0$$

where D denotes the right-hand lower derivate. Hence $p_{ij}(t)\, e^{q_j t}$ is nondecreasing in t and $D p_{ij}$ is finite a.e. Put

$$v_{ij}(t) = D p_{ij}(t) + p_{ij}(t)\, q_j \geqq 0. \tag{12}$$

Writing

$$p_{ij}(t+s)\, e^{q_j(t+s)} = e^{q_j t} \sum_k p_{ik}(t)\, p_{kj}(s)\, e^{q_j s}, \tag{13}$$

applying D with respect to s and using Fubini's theorem quoted in the preceding proof, we obtain

$$v_{ij}(t+s) = \sum_k p_{ik}(t)\, v_{kj}(s) \tag{14}$$

for each t and a.a.s. The corresponding inequality with "$\geqq$" replacing "$=$" is true for all t and s by Fatou's lemma. In particular

$$\infty > v_{ij}(t_0) \geqq p_{ii}(t_0 - s)\, v_{ij}(s)$$

for a.a. t_0 and $s < t$, from which it follows that each v_{ij} is finite everywhere. By Fubini's theorem on product measure, (14) is valid if $s \notin Z$ and $t \notin Z_s$ where $\mu(Z) = \mu(Z_s) = 0$. For an $s_0 \notin Z$ suppose that for a certain t we have

$$v_{ij}(t+s_0) > \sum_k p_{ik}(t)\, v_{kj}(s_0). \tag{15}$$

Then if $t' > t$ it follows that

$$\begin{aligned} v_{ij}(t'+s_0) &\geqq \sum_l p_{il}(t'-t)\, v_{lj}(t+s_0) > \sum_l p_{il}(t'-t) \sum_k p_{lk}(t)\, v_{kj}(s_0) \\ &= \sum_k p_{ik}(t')\, v_{kj}(s_0)\,. \end{aligned}$$

This is impossible by the choice of s_0 since (14) holds for a.a.t. Hence Z_s is empty if $s \notin Z$. Next, let $s > 0$ be arbitrary, $0 < s' < s$ and $s' \notin Z$; then

$$\begin{aligned} v_{ij}(t+s) = v_{ij}(t+s-s'+s') &= \sum_k p_{ik}(t+s-s')\, v_{kj}(s') \\ &= \sum_k \sum_l p_{il}(t)\, p_{lk}(s-s')\, v_{kj}(s') = \sum_l p_{il}(t)\, v_{lj}(s)\,. \end{aligned}$$

Hence Z is empty. Therefore (14) holds for all $t \geqq 0$, $s > 0$ and consequently v_{ij} is continuous in $\boldsymbol{T}^0$, by Theorem 2.3. The function p_{ij}, having a continuous Dini derivate Dp_{ij} in $\boldsymbol{T}^0$, has in fact a continuous derivative p'_{ij} there (see e.g. SAKS [1; p. 204]). The continuity in $\boldsymbol{T}$ follows as in the preceding proof. ▯

We remark that the difference between Theorem 1 and Theorem 2 lies in the fact that for a fixed i we have the assumption (1.B) for which there is no analogue for a fixed j. The proof of Theorem 2 uses only the finiteness of all p_{ij} and assumptions (1.A), (1.C) and (2.E) without the intervention of (1.B). Under the same assumptions the first sentence in Theorem 1 and the relation (5), namely the part of Theorem 1 which is the dual of Theorem 2, can be proved in a way exactly dual to the proof of Theorem 2. The details are left to the reader.

We take this occasion to discuss a simple but important extension of the notion of a transition matrix. A finite or denumerable array of functions (p_{ij}), $i, j \in \boldsymbol{I}$, defined in $\boldsymbol{T}^0$ will be called a *substochastic transition matrix* iff it satisfies the conditions (1.A), (1.C) and

$$\text{(B')} \qquad \sum_j p_{ij}(t) \leqq 1\,.$$

In other words the equation (1.B) for a transition matrix is relaxed into the inequality (B'). The substochastic transition matrix is called *measurable* iff it satisfies (1.D), and *standard* iff it satisfies (2.E). Most properties of a substochastic transition matrix can be derived immediately from the corresponding ones for a transition matrix, by means of the following theorem.

Theorem 3. Let (p_{ij}), $i,j \in \boldsymbol{I}$, be a substochastic transition matrix. There exists a transition matrix $(\hat{p}_{ij})$, $i,j \in \hat{\boldsymbol{I}}$, where $\hat{\boldsymbol{I}} = \boldsymbol{I} \cup \{\vartheta\}$ and ϑ is an added new index such that

$$\hat{p}_{ij} = p_{ij} \quad \text{if } i,j \in \boldsymbol{I}.$$

This $(\hat{p}_{ij})$ is measurable if and only if (p_{ij}) is.

Proof. Define $(\hat{p}_{ij})$ as follows:

$$\left.\begin{aligned}\hat{p}_{ij}(t)\equiv p_{ij}(t); \quad \hat{p}_{\vartheta j}(t)\equiv \delta_{\vartheta j};\\ \hat{p}_{i\vartheta}(t)\equiv 1-\sum_{j\in\boldsymbol{I}} p_{ij}(t);\end{aligned}\right\} \quad i, j\in\boldsymbol{I}.$$

It is a simple matter to verify that the enlarged matrix is a transition matrix. In order to prove the last assertion of the theorem we prove that $\hat{p}_{i\vartheta}$ is nondecreasing. Observe that if $s<t$,

$$\sum_j p_{ij}(t)=\sum_k p_{ik}(s)\sum_j p_{kj}(t-s)\leqq \sum_k p_{ik}(s)$$

by (B′). Hence $\sum_j p_{ij}(t)$ is nonincreasing and $\hat{p}_{i\vartheta}(t)$ is nondecreasing in t. ▯

It follows from Theorem 3 that for a substochastic transition as well as a transition matrix: measurability, continuity, and the existence of limits at zero (of all p_{ij}) are equivalent properties. Furthermore it is clear from Theorem 1.2 that in the classification of $\boldsymbol{I}$ the new index ϑ is not in F and forms a class by itself, since $\hat{u}_{\vartheta j}=\delta_{\vartheta j}$. Hence if $i\notin F$ then $\hat{u}_{i\vartheta}=0$ namely $\lim_{t\to 0}\sum_{j\in\boldsymbol{I}} p_{ij}(t)=1$. The portion of Theorem 1.3 which does not involve an index in F remains valid for a substochastic transition matrix; so do Theorems 2.4, 2.5, 2.6, 3.1 and 3.2 except that equation (3) in Theorem 3.1 should be replaced by the corresponding inequality $\sum_j p'_{ij}(t)\leqq 0$, since $\hat{p}'_{i\vartheta}(t)\geqq 0$.

Notes. Theorems 1 and 2, except the continuity of p'_{ij} at zero, were first proved by D. G. AUSTIN [1], [2] by purely analytic means. A new proof of Theorem 1 and a partial proof of (11) in Theorem 2 are given in CHUNG [5], in which probabilistic methods are used to introduce the nonnegative quantities r_{ij} and v_{ij} and to derive their properties; see §§ 15 to 16. Another proof is given by YUŠKEVIČ [2] who also gave an example in which p_{ij} does not have a finite second derivative. JURKAT [1] simplified and strengthened the ideas in CHUNG [5] and the present proofs are cast largely in his form, with a further simplification suggested by G. E. H. REUTER. A simplified version of AUSTIN'S proof of Theorem 1 together with its connection with semigroup theory is given by REUTER [2]. A substochastic transition matrix is called a (Markov) *process*, and a transition matrix an *honest process* in the last cited paper.

For simpler new analytic proofs of Theorems 1 and 2, obtained after the first edition and based on Theorem 12.4 below, see CHUNG [9]. For differentiability of transition functions in Euclidean space, see HSU [1].

§ 4. Definitions and measure-theoretic foundations

For general definitions, conventions and notation we refer to § I.1 and § II.1.

A continuous parameter stochastic process is a one-parameter family of random variables $\{x_t\}$ on a probability triple $(\Omega, \mathscr{F}, \boldsymbol{P})$. The parameter t may range over an arbitrary linear set S, although we shall be mainly concerned with the sets $\boldsymbol{T}=[0, \infty)$ and $\boldsymbol{T}^0=(0, \infty)$. For a fixed ω, the function $x(\cdot, \omega)=x.(\omega)$ of $t\in S$ is called the *sample function corresponding to* ω. Its domain is S; let its range be $\mathscr{R}(\omega)$. The union of $\mathscr{R}(\omega)$ for all $\omega\in\Omega$ is the *range* of the process, but since a null set is usually immaterial the more pertinent concept is that of the *essential range*, defined to be the set of values each of which is contained in $\mathscr{R}(\omega)$ for a set of ω of positive probability. The essential range of a process is therefore invariant if the sample functions of a null set are changed. We shall be mainly concerned with a process in which each x_t is a discrete random variable and moreover the union $\boldsymbol{I}$ of all the possible values (see § I.1) of all the x_t is still a denumerable set. Such a process will be called *denumerably-valued* and $\boldsymbol{I}$ its *minimal state space*. In contrast to the discrete parameter case, the essential range of a denumerably-valued process may be larger than the minimal state space; in fact it may be non-denumerable. A value in the essential range will be called a *state* of the process; and it will be called a *fictitious state* iff it is not in $\boldsymbol{I}$. As a convention an unspecified state shall not be fictitious.

A continuous parameter Markov chain is a stochastic process $\{x_t, t\in\boldsymbol{T}\}$ which is denumerably-valued with the minimal state space $\boldsymbol{I}$ and which possesses the *Markov property*: for every $n\geqq 2$, $0\leqq t_1<\cdots<t_n$ and any $i_1, \ldots, i_n$ in $\boldsymbol{I}$ we have

$$(1)\qquad \begin{cases} \boldsymbol{P}\{x_{t_n}(\omega)=i_n \mid x_{t_1}(\omega)=i_1, \ldots, x_{t_{n-1}}(\omega)=i_{n-1}\} \\ \qquad\qquad\qquad =\boldsymbol{P}\{x_{t_n}(\omega)=i_n \mid x_{t_{n-1}}(\omega)=i_{n-1}\}\end{cases}$$

whenever the left member is defined. This is the same condition as (I.1.1) except that here the t_ν's are real numbers instead of integers. As in the discrete parameter case, the Markov property implies the following: if $\Lambda\in\mathscr{F}\{x_s, s\leqq t\}$, and $\mathsf{M}\in\mathscr{F}\{x_s, s>t\}$ then

$$\boldsymbol{P}\{\mathsf{M}\mid\Lambda; x_t\}=\boldsymbol{P}\{\mathsf{M}\mid x_t\}$$

with probability one. The chain is said to have *stationary transition probabilities* or to be *homogeneous* iff the conditional probability $\boldsymbol{P}\{x_{s+t}(\omega)=j \mid x_s(\omega)=i\}$ does not depend on s whenever defined. Such a process will be abbreviated as a "M.C.", prefixed by "c.p." when the distinction between the continuous parameter case and the discrete parameter ("d.p.") case of Part I is stressed. Now the preceding conditional probability is defined for some $s \geqq 0$ if and only if $i \in \boldsymbol{I}$ by definition of the minimal state space. Hence for $i, j \in \boldsymbol{I}$ we may put

$$(2) \qquad p_{ij}(t) = \boldsymbol{P}\{x_{s+t}(\omega) = j \mid x_s(\omega) = i\}, \qquad t > 0;$$

namely, the left member is the common value of the right member for all s for which the latter is defined. The function p_{ij} is called the *transition probability function from the state i to the state j*: its value $p_{ij}(t)$ at t the *transition probability from i to j in time t*. The *absolute distribution at time $t \geqq 0$* is defined to be $\{p_i(t), i \in \boldsymbol{I}\}$ where

$$(3) \qquad p_i(t) = \boldsymbol{P}\{x_t(\omega) = i\}, \qquad \sum_i p_i(t) = 1.$$

In particular, $p_i(0)$ is abbreviated to p_i (although this conflicts with the notation p_i as the function $p_i(\cdot)$):

$$(4) \qquad p_i = \boldsymbol{P}\{x_0(\omega) = i\}, \qquad \sum_i p_i = 1;$$

and $\{p_i, i \in \boldsymbol{I}\}$ is called the *initial distribution*. We have the obvious relation

$$(5) \qquad p_j(t) = \sum_i p_i p_{ij}(t), \qquad t \geqq 0.$$

The probability distribution of finite collections of the x_t's are called the *(finite-dimensional) joint distributions* of the process. If $0 \leqq t_0 < t_1 < \cdots < t_n$, we have by the properties of conditional probabilities and repeated use of the Markov property:

$$(6) \qquad \boldsymbol{P}\{x_{t_\nu}(\omega) = i_\nu, 1 \leqq \nu \leqq n \mid x_{t_0}(\omega) = i_0\} = \prod_{\nu=1}^{n} p_{i_{\nu-1} i_\nu}(t_\nu - t_{\nu-1}).$$

A particular case of this may be written

$$\boldsymbol{P}\{x_{t_0+s}(\omega) = k, x_{t_0+s+t}(\omega) = j \mid x_{t_0}(\omega) = i\} = p_{ik}(s) p_{kj}(t);$$

and summing over $k \in \boldsymbol{I}$ we obtain (1.C). Since (1.A) and (1.B) are trivial we have shown that the matrix (p_{ij}), $i, j \in \boldsymbol{I}$ of functions on $\boldsymbol{T}^0$ defined by (2) is a transition matrix in the sense of § 1. Henceforth it will be called the *transition matrix of the* M.C.; its index set is the minimal state space $\boldsymbol{I}$.

Conversely, given a transition matrix (p_{ij}), $i, j \in \boldsymbol{I}$, in the sense of § 1 and an arbitrary distribution $\{p_i, i \in \boldsymbol{I}\}$, there exists a M.C. $\{x_t, t \in \boldsymbol{T}\}$ with the minimal state space $\boldsymbol{I}$, for which (2) and (4), and consequently also (6) hold. This is the analogue of Theorem I.2.1; the proof is also analogous and will be only briefly sketched. The space Ω is the space of all functions with domain $\boldsymbol{T}$ and range $\boldsymbol{I}$; each ω is such a function $\xi(\cdot)$. The Borel field $\mathscr{F}$ is the smallest Borel field containing all sets of the form $\{\xi(\cdot) : \xi(t) = i\}$, $t \in \boldsymbol{T}$, $i \in \boldsymbol{I}$. The probability measure $\boldsymbol{P}$ on the function space Ω is defined as follows: if $0=t_0<t_1<\cdots<t_n$, the set of functions $\xi(\cdot)$ such that $\xi(t_\nu)=i_\nu$ for $0\leqq\nu\leqq n$ is assigned the measure

$$p_{i_0}\prod_{\nu=1}^{n} p_{i_{\nu-1}i_\nu}(t_\nu-t_{\nu-1}).$$

These assignments of joint distributions are mutually consistent because of the properties of a transition matrix. Hence the measure $\boldsymbol{P}$ can be extended to the entire $\mathscr{F}$ by KOLMOGOROV's theorem. The Markov process to be constructed is then defined as follows: $x_t(\omega)=\xi(t)$ if ω is the function $\xi(\cdot)$; in other words the sample function corresponding to $\omega=\xi(\cdot)$ is precisely $\xi(\cdot)$.

The preceding discussion for $\{x_t,\ t\in\boldsymbol{T}\}$ needs only obvious changes for $\{x_t,\ t\in\boldsymbol{T}^0\}$. The defining property (1) is to hold only for $0<t_1$. There is no longer an initial distribution and the joint distributions must be determined by the transition matrix and the absolute distribution $\{p_i(t),\ i\in\boldsymbol{I}\}$ for t in an arbitrary interval $(0,\varepsilon)$ or at least for a sequence of values of t tending to zero. There must be consistency among the given $\{p_i(\cdot)\}$ such that if it is given at both s and t, $s<t$, then we have

$$p_j(t)=\sum_i p_i(s)\,p_{ij}(t-s). \tag{7}$$

If $\{p_i(\varepsilon)\}$ is given then if $\varepsilon<t_1<\cdots<t_n$ we have

$$\boldsymbol{P}\{x_{t_\nu}(\omega)=i_\nu,\ 1\leqq\nu\leqq n\}=\sum_{i_0}p_{i_0}(\varepsilon)\prod_{\nu=1}^{n}p_{i_{\nu-1}i_\nu}(t_\nu-t_{\nu-1})$$

where $t_0=\varepsilon$. We shall first encounter this kind of process in §9. Unless otherwise stated, our discussion is concerned with a M.C. in $\boldsymbol{T}$.

It is well known that a stochastic process defined through given joint distributions, as in the construction above, is still inadequate for many purposes. Further specifications must be made before certain problems become meaningful. We proceed to review the relevant parts of the general theory for a general process on an arbitrary linear parameter set $\boldsymbol{T}$, not necessarily $[0,\infty)$.

Two arbitrary stochastic processes $\{x_t\}$ and $\{\tilde{x}_t\}$ defined on the same probability triple and parameter set $\boldsymbol{T}$ are called *standard modifications* or *versions* of each other iff

$$\boldsymbol{P}\{x_t(\omega)=\tilde{x}_t(\omega)\}=1 \tag{8}$$

for every fixed $t\in\boldsymbol{T}$. It follows that if R is any denumerable subset of $\boldsymbol{T}$, then

$$\boldsymbol{P}\{x_t(\omega)=\tilde{x}_t(\omega)\ \text{ for all }\ t\in R\}=1.$$

In particular, corresponding joint distributions are identical, since

$$\boldsymbol{P}\{x_t(\omega)\in A_t,\ t\in R\}=\boldsymbol{P}\{\tilde{x}_t(\omega)\in A_t,\ t\in R\} \tag{9}$$

where each A_t is a linear Borel set. For a nondenumerable set R however the equation (9) need not hold. Indeed, it may happen that of the two ω-sets in question, one set belongs to $\mathscr{F}$ while the other does not, so that one member of (9) is defined while the other is not. Since ω-sets which describe the analytical properties of sample functions are of primary importance in the study of the process we are led to seek a version in which such sets will belong to $\mathscr{F}$. We begin with the basic concepts of separability and measurability. In the following the real line in the Euclidean topology is compactified by the two points at infinity $\pm\infty$; thus a "closed linear set" refers to a closed subset of the extended line $[-\infty, \infty]$.

The process $\{x_t, t\in \boldsymbol{T}\}$ is said to be *separable (relative to the class of closed sets)* iff there exists a denumerable subset R of $\boldsymbol{T}$ and a null set N with the following property: for any closed linear set A and any open interval G in $(-\infty, \infty)$ we have

$$(10) \qquad \{\omega: x_t(\omega)\in A, t\in G\cap R\} - \{\omega: x_t(\omega)\in A, t\in G\cap \boldsymbol{T}\} \subset \mathsf{N}.$$

The set R is called a *separability set*. Since a denumerable subset of $\boldsymbol{T}$ containing R is a separability set we may always suppose that R is dense in $\boldsymbol{T}$. The process is said to be *well-separable* if any denumerable subset dense in $\boldsymbol{T}$ is a separability set.

In order to describe the sample functions of a separable process let us make the following definitions. A real and possibly infinite-valued function f in $\boldsymbol{T}$ is *separable* at t with respect to a set R dense in $\boldsymbol{T}$ iff the value $f(t)$ is a limiting value of $f(r)$, $r\neq t$, $r\in R\cap(t-n^{-1}, t+n^{-1})$, for every n; f is separable in $S\subset \boldsymbol{T}$ iff it is separable at every $t\in S$. We recall also that the process is said to be *stochastically continuous* at $t_0\in \boldsymbol{T}$ iff t_0 is in the closure of $\boldsymbol{T}$ and $x_t\to x_{t_0}$ in probability, namely $\lim_{t\to t_0} \boldsymbol{P}\{|x_t(\omega)-x_{t_0}(\omega)|>\varepsilon\}=0$ for every ε; it is said to be stochastically continuous in $S\subset \boldsymbol{T}$ iff it is so at every $t_0\in S$, and simply stochasticaliy continuous when $S=\boldsymbol{T}$. The following theorem is a restatement of the separability definition in terms of sample functions, with an important strengthening. It will be referred to in the sequel by the phrase: "by separability".

Theorem 1. The process $\{x_t, t\in \boldsymbol{T}\}$ is separable with R as a separability set if and only if almost all sample functions are separable in $\boldsymbol{T}-R$. If so and if the process is stochastically continuous, then almost all sample functions are separable in $\boldsymbol{T}$.

Proof. Suppose the process is separable with R. Let A_n denote the closure of values $x(r,\omega)$, $|r-t|<n^{-1}$, $r\in R\cap \boldsymbol{T}$; and let $A=\bigcap_{n=1}^{\infty} A_n$. It

follows from (10) that if $\omega \notin \mathsf{N}$ and $t \notin R$ then $x(t, \omega) \in A_n$ for every n and consequently $x(t, \omega) \in A$. Thus $x(\cdot, \omega)$ is separable in $\boldsymbol{T}-R$. The converse is similarly proved. Next, if the process is stochastically continuous at $r_0 \in R$, there exists a sequence $\{t_m\}$ in $\boldsymbol{T}-R$ tending to r_0 such that $\lim_{m\to\infty} x(t_m, \omega) = x(r_0, \omega)$ for a.a. ω. Hence $x(r_0, \omega)$ lies in the closure of values $x(t_m, \omega)$, $|t_m - r_0| < n^{-1}$. If the process is also separable with R then by (10) each $x(t_m, \omega)$ lies in turn in the closure of values $x(r, \omega)$, $r \in R$, $|r - t_m| < n^{-1}$, provided $\omega \notin \mathsf{N}$. This being true for all n, $x(\cdot, \omega)$ is separable at r_0 for a.a. ω. Since R is denumerable, this implies that $x(\cdot, \omega)$ is separable in R for a.a. ω. Combinging this result with the first assertion of the theorem we obtain the second. ▯

A fundamental theorem due to Doob [4; Theorem II.2.4 and pp. 59 to 60] states that any stochastic process has a separable version. The set R may be so chosen that the essential range of this version is contained in the closure of the essential range of the original process. In particular the new sample functions may assume infinite values even if the original ones do not. To return to our M.C., this is where the fictitious states enter the picture.

Theorem 2. Let the M.C. $\{x_t, t \in \boldsymbol{T}\}$ be separable. If y is not in the closure $\bar{\boldsymbol{I}}$ of $\boldsymbol{I}$ (in the Euclidean topology), then we have

$$(11) \qquad \boldsymbol{P}\{x(t, \omega) \neq y \text{ for all } t \in \boldsymbol{T}\} = 1;$$

in other words y is not in the essential range of the process.

Proof. Since $y \notin \bar{\boldsymbol{I}}$ we have $\boldsymbol{P}\{x(r, \omega) \neq y\} = 1$ for every $r \in R$; since R is denumerable this implies that

$$(12) \qquad \boldsymbol{P}\{x(r, \omega) \neq y \text{ for all } r \in R\} = 1.$$

By hypothesis $\bar{\boldsymbol{I}} - \{y\}$ is a closed set. Taking this set to be A and any open interval containing $\boldsymbol{T}$ to be G in (10), we see that (12) implies (11) on account of separability. ▯

Theorem 2 shows that the essential range of a separable M.C. depends on the topological nature of $\boldsymbol{I}$. A complete theory should therefore allow the most general denumerable set $\boldsymbol{I}$, indeed in an abstract space with an arbitrary topology rather than the Euclidean real line considered here. However, the theory of M.C. has not yet been sufficiently developed to make this generality feasible, and we must be content with the simplest possible situation, namely that of a *discrete set* $\boldsymbol{I}$ with the *discrete topology* (or, if we confine ourselves to the real line since we have defined a random variable to be real-valued, an *isolated* set of real numbers). This set, if infinite, will be compactified by the well known one-point compactification (see e.g. LOÈVE [1; p. 71]) as follows. We adjoin a new point ∞ to $\boldsymbol{I}$, put $\boldsymbol{I} \cup \{\infty\} = \bar{\boldsymbol{I}}$, and postulate that every infinite subset of $\bar{\boldsymbol{I}}$ is closed if and only if it contain ∞. The enlarged set $\bar{\boldsymbol{I}}$ will be called the *compactified (minimal) state space*. According to Theorem 2 then, a separable M.C. with such a minimal state space $\boldsymbol{I}$ either has no fictitious state at all or has the unique one ∞. The former case arises of course whenever $\boldsymbol{I}$ is finite, but it can also arise when $\boldsymbol{I}$ is infinite as in the Poisson process or more generally when almost all sample functions are step functions (see §19). The latter case is more interesting and will be amply illustrated in the sequel (see in particular §20); in this case ∞ will be called the *adjoined state*. We may enlarge the index set for the absolute distribution and the transition matrix to $\bar{\boldsymbol{I}}$ if we put

$$(13) \qquad p_{\infty}(t) \equiv 0, \quad p_{i\infty}(t) \equiv 0, \quad p_{\infty i}(t) \equiv \delta_{\infty i}, \qquad i \in \boldsymbol{I},\ t \in \boldsymbol{T}.$$

The last definition is of course purely formal. The enlarged matrix (p_{ij}), $i, j \in \bar{\boldsymbol{I}}$, is a transition matrix and equations (2) to (5) above continue to hold with the summation in (5) taken over $\boldsymbol{I}$ or $\bar{\boldsymbol{I}}$ indifferently. Finally we make the convention that $\infty > i$ for each $i \in \boldsymbol{I}$ as if $\boldsymbol{I}$ were the set of positive integers.

We now proceed to the concept of measurability of a general process $\{x_t, t \in \boldsymbol{T}\}$. The following discussion is valid if the parameter set is an arbitrary linear Borel set. Consider the product space $\boldsymbol{T} \times \Omega$ with the generic point (t, ω). Let $\mathscr{B}$ denote the usual Borel field over $\boldsymbol{T}$, namely the smallest Borel field containing all (open) intervals (a set in $\mathscr{B}$ being a linear Borel set). On $\mathscr{B}$ let μ be the usual Borel-Lebesgue measure. Observe that the couple $(\mathscr{B}, \mu)$ is not complete, its completion being the Borel field of Lebesgue measurable sets with the measure μ. Now consider the product Borel field $\mathscr{B} \times \mathscr{F}$, with the product measure $\mu \times \boldsymbol{P}$. The *completion* of the couple $(\mathscr{B} \times \mathscr{F}, \mu \times \boldsymbol{P})$ is the smallest Borel field containing $\mathscr{B} \times \mathscr{F}$ and all subsets of sets in $\mathscr{B} \times \mathscr{F}$ having $\mu \times \boldsymbol{P}$ measure zero; it will be denoted by $\overline{\mathscr{B} \times \mathscr{F}}$ and the extension of $\mu \times \boldsymbol{P}$ to the enlarged field will be denoted by the same symbol. "Almost all" or "a.a." (t, ω) will refer to this measure. A set of (t, ω) is called measurable or Borel measurable according as it belongs to $\overline{\mathscr{B} \times \mathscr{F}}$ or $\mathscr{B} \times \mathscr{F}$; similarly for a function of (t, ω).

The stochastic process $\{x_t, t \in \boldsymbol{T}\}$ is called *measurable* or *Borel measurable* according as $x(\cdot, \cdot)$ as a function of (t, ω) on $\boldsymbol{T} \times \Omega$ is measurable or Borel measurable. In the remainder of this section only the less stringent notion of measurability is discussed; Borel measurability will be taken up in §7.

If the process $\{x_t\}$ is separable and stochastically continuous, then it is well-separable (DOOB [4; Theorem II.2.2(i) and pp. 59 to 60]). Moreover, in this case it has a standard modification which is well-separable and measurable (Doob [4; Theorem II.2.6]). Specializing to our M.C., the following theorem links its separability and measurability to its transition matrix. It is not assumed that $\boldsymbol{I}$ is a discrete set.

Theorem 3. Suppose that the M.C. $\{x_t, t\in \boldsymbol{T}\}$ has a measurable transition matrix such that the set F of indices in Theorem 1.2 and 1.3 is empty; or equivalently for every $i\in\boldsymbol{I}$ we have $p_{ii}(t)\neq 0$ in $\boldsymbol{T}^0$. Then it has a well-separable and measurable standard modification if and only if it is stochastically continuous, or equivalently if and only if its transition matrix is standard.

Proof. We shall prove that a standard transition matrix implies stochastic continuity, that the latter implies a well-separable and measurable version, which in turn implies a standard transition matrix. Let us assume then that the transition matrix is standard. If $s<t$, we have

$$\boldsymbol{P}\{x_s(\omega)=x_t(\omega)\}=\sum_j p_j(s)\,p_{jj}(t-s);$$

consequently by Fatou's lemma, and the continuity of $p_j(\cdot)$,

$$\varliminf_{s\uparrow t} \boldsymbol{P}\{x_s(\omega)=x_t(\omega)\}\geqq \sum_j p_j(t)=1.$$

Hence the limit exists and is equal to one. If $t<s$, we have

$$\boldsymbol{P}\{x_s(\omega)=x_t(\omega)\}=\sum_j p_j(t)\,p_{jj}(s-t);$$

consequently by uniform convergence the limit as $s\downarrow t$ is equal to one. Therefore the process is stochastically continuous and there is a well-separable and measurable version by the theorem quoted above.

Conversely if there is such a version then since the transition matrix remains the same we may as well suppose that the original process is separable and measurable. For a fixed $s>0$ and each $h>0$ put

$$\Gamma_h=\{(t,\omega): |x(t+h,\omega)-x(t,\omega)|>\varepsilon\};$$

$$c_h(t,\omega)=\begin{cases}1 & \text{if } (t,\omega)\in\Gamma_h,\\ 0 & \text{otherwise};\end{cases}$$

so that $c_h(\cdot,\cdot)$ is the *indicator* of the set Γ_h. Since the process is measurable we have $\Gamma_h\in\overline{\mathscr{B}\times\mathscr{F}}$. Let the *sections* of Γ_h determined by fixing ω and t respectively be

$$A_h(\omega)=\{t:\ (t,\omega)\in\Gamma_h\}\quad\text{and}\quad \Lambda_h(t)=\{\omega:\ (t,\omega)\in\Gamma_h\}.$$

Since $x(\cdot,\cdot)$ is measurable $\overline{\mathscr{B}\times\mathscr{F}}$, $x(\cdot,\omega)$ is measurable $\overline{\mathscr{B}}$ (i.e. Lebesgue measurable) for a.a. ω by Fubini's theorem; hence by a well known theorem (a special case is cited in the proof of Theorem 1.1), $x(t+h,\omega)-$

$x(t,\omega)$ converges to zero in μ-measure in every finite interval $T_1 \subset \boldsymbol{T}$ as $h \to 0$, that is

$$\lim_{h\to 0} \mu[T_1 \cap A_h(\omega)] = 0$$

for a.a. ω. It follows that

$$\lim_{h\to 0} \int_\Omega \int_{T_1} c_h(t,\omega)\, dt\, \boldsymbol{P}(d\omega) = \lim_{h\to 0} \int_\Omega \mu[T_1 \cap A_h(\omega)]\, \boldsymbol{P}(d\omega) = 0.$$

Inverting the double integral by Fubini's theorem, we obtain

$$\lim_{h\to 0} \int_{T_1} \boldsymbol{P}[\Lambda_h(t)]\, dt = \lim_{h\to 0} \int_{T_1} \int_\Omega c_h(t,\omega)\, \boldsymbol{P}(d\omega)\, dt = 0.$$

Hence the function $\boldsymbol{P}[\Lambda_h(\cdot)]$ converges to zero in μ-measure in the interval T_1, and unless $t \in Z \subset T_1$ where $\mu(Z) = 0$, there exists a sequence $\{h_n\}$ decreasing to zero such that $\lim_{n\to\infty} \boldsymbol{P}[\Lambda_{h_n}(t)] = 0$.

Now let i and j be two distinct elements of $\boldsymbol{I}$ and choose $\varepsilon < |i-j|$. If $p_i(t) \equiv 0$ in $\boldsymbol{T}^0$ then since $i \in \boldsymbol{I}$ we must have $p_i > 0$. Consequently $p_i p_{ii}(t) \leqq p_i(t)$ and so $p_{ii}(t) \equiv 0$ in $\boldsymbol{T}^0$ which contradicts the assumption that F is empty. Hence $p_i(t_0) > 0$ for some $t_0 \in \boldsymbol{T}^0$. Since the transition matrix is measurable, it follows from Theorem 1.1 and (5) above that $p_i(t)$ is continuous in t. Choose T_1 to contain t_0; then there exists a $t_1 \in T_1 - Z$ such that $t_1 > t_0$ and $p_i(t_1) > 0$. We have from the above

$$\begin{aligned} 0 = \lim_{n\to\infty} \boldsymbol{P}[\Lambda_{h_n}(t_1)] &= \lim_{n\to\infty} \sum_k \boldsymbol{P}\{x(t_1,\omega) = k;\ |x(t_1+h_n,\omega) - k| > \varepsilon\} \\ &\geqq \lim_{n\to\infty} p_i(t_1)\, p_{ij}(h_n). \end{aligned}$$

It follows that $\lim_{n\to\infty} p_{ij}(h_n) = 0$. By Theorem 1.3 and using the notation there, this means $u_{ij} = 0$. Hence $u_{ii} = \lim_{t\to 0} p_{ii}(t) = 1$ by (1.9) and (2.E) is satisfied. ▯

Remark. The separability and measurability of a M.C. do not imply that its transition matrix is standard, if F is not empty. Example: $p_1 = 1$, $p_2 = 0$, $p_{11}(t) \equiv p_{21}(t) \equiv 0$, $p_{12}(t) \equiv p_{22}(t) \equiv 1$; for all $\omega \in \Omega$, let $x(0,\omega) = 1$; $x(t,\omega) \equiv 2$, $t \in \boldsymbol{T}^0$. The point is that the initial distribution can be chosen independently of the transition matrix; hence an index in F may still correspond to a nonfictitious state.

Notes. Much of this section is of course pedagogy and semantics. The term "fictitious state" was introduced by Lévy [1] but the term "state" does not seem to have been precisely defined before. Is a fictitious state a state and can we arbitrarily add a (fictitious) state? The answers are given by the definitions proposed here, without claim to sanctimony. It is customary to take the minimal state space to be the set of positive, nonnegative, or all integers. In the last case it is more natural, but by no means mandatory, to adjoin two states $+\infty$ and $-\infty$. The role of fictitious states is clearly recognized by LÉVY although he did not mention separability. Even when we assume $\boldsymbol{I}$ to be discrete and compactify it by adjoining one fictitious state, as we shall do from here on in the monograph, this single adjoined state plays an essential role in all rigorous discussions of sample functions (unless severe assumptions are imposed on the process; see e.g. those of Theorem 19.2). Our discussion concerning a general denumerable set $\boldsymbol{I}$, while not necessary here, should serve as introductory remarks to a more complete theory of compactification of the minimal state space. In the same vein Theorem 3 is given in a slightly extended form over that of DOOB [1].

It should be stressed that separability in this monograph means separability relative to the class of closed sets rather than closed intervals as the term is used in DOOB [4]. The latter weaker concept is not adequate here.

§ 5. The sets of constancy

From now on we assume that the minimal state space $\boldsymbol{I}$ of the M.C. $\{x_t, t\in\boldsymbol{T}\}$ is discrete and that it is compactified to $\bar{\boldsymbol{I}}$ by adjoining the fictitious state ∞. We also assume that the transition matrix (p_{ij}) is standard. Instead of going at once to a well-separable and measurable version as guaranteed by Theorem 4.3, we shall first investigate separately the consequences of separability and measurability. In this and the next two sections a theorem which is proved under the further assumption of separability is marked with (S), that of measurability with (M), that of both with (SM).

As a general notation which will be used throughout the sequel, we write for any $i\in\bar{\boldsymbol{I}}$,

$$S_i(\omega)=\{t\colon x(t,\omega)=i\}$$

so that $\{\omega\colon t\in S_i(\omega)\}=\{\omega\colon x(t,\omega)=i\}$; and we write $\overline{S_i(\omega)}$ for the closure of $S_i(\omega)$. We call $S_i(\omega)$ the *i-set of the sample function* $x(\cdot,\omega)$ or a *set of constancy* when i is not specified. If $\{x_t\}$ is a measurable process, then $S_i(\omega)$ is a measurable t-set for a.a. ω by FUBINI's theorem. Let $\xi(\cdot,\omega)$ denote its indicator:

$$\xi(t,\omega)=\begin{cases}1 & \text{if } t\in S_i(\omega),\\ 0 & \text{otherwise.}\end{cases}$$

Then $\xi(\cdot,\cdot)$ is measurable in (t,ω) and for every measurable $T_1\subset\boldsymbol{T}$

$$\int_{T_1}\xi(t,\omega)\,dt=\mu[S_i(\omega)\cap T_1]$$

is a random variable.

Theorem 1 (M). We have

$$P\{\mu[S_i(\omega)]=0\}=1$$

if and only if $i=\infty$.

Proof. By Fubini's theorem,

$$E\{\mu[S_i(\omega)]\}=\int_0^\infty p_i(t)\,dt. \tag{1}$$

If $i=\infty$ then $p_i(t)\equiv 0$ in $\boldsymbol{T}$; if $i\neq\infty$ then $p_i(t)\not\equiv 0$ in $\boldsymbol{T}$ and is continuous in $\boldsymbol{T}$. Hence the right member of (1) is equal to zero if and only if $i=\infty$; the theorem follows. ▯

It is possible that $P\{\overline{S_\infty(\omega)}=\boldsymbol{T}\}=1$; see Example 20.6.

Before we proceed further we make an interesting digression. The following new proof of Theorem 1.5 depends essentially on the separability and measurability of an associated M.C.

Theorem 2. Let (p_{ij}), $i,j\in\boldsymbol{I}$, be a measurable transition matrix. Then each p_{ij} is either identically zero or never zero in $\boldsymbol{T}^0$.

Proof. By Theorem 1.3 it is sufficient to prove the theorem for a standard transition matrix. Let j be arbitrary but fixed; then by Theorem 1.4 for each i there exists a $u(i)$, $0\leqq u(i)\leqq\infty$, such that $p_{ij}(t)=0$ if and only if $t\in(0,u(i)]$. If $u(i)=0$ this interval is empty while if $u(i)=\infty$ it should be $\boldsymbol{T}^0$. Theorem 2 asserts that for each j, $u(i)$ is either ∞ or 0. We know from Theorem 1.4 that $u(j)=0$. Let i be fixed, $i\neq j$.

Let $\{x_t, t\in\boldsymbol{T}\}$ be a well-separable, measurable M.C. with the compactified state space $\overline{\boldsymbol{I}}$, the given transition matrix, and the initial distribution concentrated at i. Let R be a denumerable dense set in $\boldsymbol{T}^0$. We may choose the probability space Ω so that for every $\omega\in\Omega$: (i) $x(0,\omega)=i$; (ii) $x(\cdot,\omega)$ is separable with respect to R in $\boldsymbol{T}$; (iii) $\mu[S_\infty(\omega)]=0$; (iv) $R\cap S_\infty(\omega)=0$. The conditions (ii) and (iii) are possible by Theorems 4.1 and 5.1 respectively. Furthermore, it follows from the definition of $u(\cdot)$ that Ω may be chosen so as to have also the following property: (v) for any $k\in\boldsymbol{I}$ and $r\in R\cap S_k(\omega)$ we have $(r, r+u(k))\cap S_j(\omega)=0$; on the other hand, given any $\varepsilon>0$, any non-null subset Λ_1 of $\{\omega: x(r,\omega)=k\}$ belonging to $\mathscr{F}\{x_t, t\leqq r\}$ contains a non-null subset Λ_2 such that $(r, r+u(k)+\varepsilon)\cap S_j(\omega)\neq 0$ for all $\omega\in\Lambda_2$.

Suppose now that $0<u(i)<\infty$ and we shall derive a contradiction. Denote the interval $[0,u(i)]$ by V and consider $\{y_t, t\in V\}$ where

$$y(t,\omega)=t+u\bigl(x(t,\omega)\bigr).$$

For each $r\in R$, $y(r,\omega)$ is defined but may be ∞. We prove that for each $\omega\in\Omega$, $y(r,\omega)$ is nondecreasing in $r\in R$. For otherwise there exist

$r, r' \in R \cap V$, $r < r'$, $e > 0$, $k, l \in \boldsymbol{I}$ such that the set $\Lambda = \{\omega : x(r, \omega) = k, x(r', \omega) = l\}$ is non-null and if $\omega \in \Lambda$,

$$r + u(k) = y(r, \omega) \geqq y(r', \omega) + \varepsilon = r' + u(l) + \varepsilon .$$

Hence $(r, r + u(k)) \subset (r', r' + u(l) + \varepsilon)$, which is impossible since by (v) above, for ω belonging to a certain non-null subset of Λ the interval on the left is disjoint from $S_j(\omega)$ while that on the right is not.

To avoid infinite values we define $v(k) = \min(u(k), u(i))$ and put

$$z(t, \omega) = t + v(x(t, \omega)) .$$

Clearly $\{z_t, t \in V\}$ is a stochastic process and $z(t, \omega)$ is defined and finite wherever $t \notin S_\infty(\omega)$. The measurability and stochastic continuity of the z-process follow from those of the x-process and the denumerability of $\boldsymbol{I}$. By DOOB's theorem cited in §4 there is a version $\{\tilde{z}_t, t \in V\}$ of the z-process which is well-separable and measurable. Since $z(\cdot, \cdot)$ and $\tilde{z}(\cdot, \cdot)$ are both measurable $\overline{\mathscr{B} \times \mathscr{F}}$ and they are standard modifications of each other, FUBINI's theorem implies the following. There is a set $\Omega_0 \subset \Omega$ with $\boldsymbol{P}(\Omega_0) = 1$ and for each $\omega \in \Omega_0$ a subset $T_0(\omega)$ of V, containing R and disjoint from $S_\infty(\omega)$, such that $\mu[T_0(\omega)] = \mu(V)$; and if $\omega \in \Omega_0$ and $t \in T_0(\omega)$ then

$$\tilde{z}(t, \omega) = z(t, \omega) = t + v(x(t, \omega))$$

and $x(t, \omega) \in \boldsymbol{I}$. Furthermore Ω_0 may be so chosen that for each $\omega \in \Omega_0$, $z(\cdot, \omega)$ is separable with respect to R in $\boldsymbol{T}$. From now on we consider an arbitrary fixed point ω in Ω_0. Then $\tilde{z}(r, \omega) = z(r, \omega)$ is nondecreasing in $r \in R \cap V$ just as $y(r, \omega)$ is; and it follows from the separability of $\tilde{z}(\cdot, \omega)$ that $\tilde{z}(t, \omega)$ is nondecreasing in $t \in V$. Next if $t \in T_0(\omega)$, for each $\delta > 0$ there exists an $r \in R \cap V \cap (t - \delta, t + \delta)$ such that $x(r, \omega) = x(t, \omega)$ since $x(t, \omega) \in \boldsymbol{I}$ and $x(\cdot, \omega)$ is separable at t. Hence

$$\tilde{z}(r, \omega) - \tilde{z}(t, \omega) = r - t$$

and thus the right-hand upper derivate $\overline{D}^+ \tilde{z}(t, \omega) \geqq 1$. Since $\tilde{z}(\cdot, \omega)$ is nondecreasing the derivative $D\tilde{z}(\cdot, \omega)$ exists a.a. in V, say in $T_1(\omega)$. Thus $D\tilde{z}(t, \omega) \geqq 1$ if $t \in T_0(\omega) \cap T_1(\omega)$, namely a.e. in V. It follows from a well known theorem (see e.g. TITCHMARSH [1; p. 361]) that if $t \in V$,

$$\tilde{z}(t, \omega) - \tilde{z}(0, \omega) \geqq \int_0^t D\tilde{z}(s, \omega)\, ds \geqq \int_0^t 1\, ds = t .$$

Therefore we have, if $t \in T_0(\omega)$,

$$u(i) \geqq v(x(t, \omega)) \geqq v(x(0, \omega)) = v(i) = u(i) .$$

In particular for any $r_0 \in R \cap V$, we have $u(x(r_0, \omega)) \geqq u(i)$ which implies by the property (v) of Ω that $(r_0, r_0 + u(i)) \cap S_j(\omega) = 0$. Thus $(0, r_0 +$

$u(i))\cap S_j(\omega)=0$ for every $\omega\in\Omega_0$, contradicting the second part of property (v) since $r_0>0$. This contradiction disproves the assumption that $0<u(i)<\infty$ and so proves the theorem. ▯

By FUBINI's theorem, we have if $t>\varepsilon$ and $i\neq\infty$,

$$\boldsymbol{E}\left\{\frac{1}{2\varepsilon}\mu[S_i(\omega)\cap(t-\varepsilon,t+\varepsilon)]\,|\,x(t,\omega)=i\right\}=\frac{1}{2\varepsilon}\int_0^{\varepsilon}\left(\frac{p_i(t-s)}{p_i(t)}+1\right)p_{ii}(s)\,ds.$$

Hence the conditional expectation tends to one as $\varepsilon\downarrow 0$. It follows that $\mu[S_i(\omega)\cap(t-\varepsilon,t+\varepsilon)]/2\varepsilon$ tends to one in probability under the hypothesis $x(t,\omega)=i$. The following theorem strengthens this result.

Theorem 3 (M). For each fixed $t>0$, we have if $i\neq\infty$,

$$\boldsymbol{P}\left\{\lim_{\varepsilon\downarrow 0}\frac{1}{2\varepsilon}\mu[S_i(\omega)\cap(t-\varepsilon,t+\varepsilon)]=1\,|\,x(t,\omega)=i\right\}=1, \tag{2}$$

$\left(\text{For } t=0 \text{ the limit is to be replaced by} \lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\mu[S_i(\omega)\cap(0,\varepsilon)].\right)$

Proof. Let $0<\varepsilon<t$ and define $e(s,t)=0$ or 1 according as $|s-t|\geqq\varepsilon$ or $<\varepsilon$ and let ξ be as on p. 148. The function $e(s,t)\xi(s,\omega)$ of the triple (s,t,ω) is then measurable $\overline{\mathscr{B}\times\mathscr{B}\times\mathscr{F}}$. Hence by FUBINI's theorem,

$$\mu[S_i(\omega)\cap(t-\varepsilon,t+\varepsilon)]=\int_0^{\infty}e(s,t)\xi(s,\omega)\,ds$$

is measurable $\overline{\mathscr{B}\times\mathscr{F}}$. Thus for each ε, $\mu[S_i(\omega)\cap(t-\varepsilon,t+\varepsilon)]$ as a function of (t,ω) is measurable $\overline{\mathscr{B}\times\mathscr{F}}$ and consequently by letting $\varepsilon_n=n^{-1}$ and $n\to\infty$ we see that

$$\Gamma=\left\{(t,\omega):\lim_{\varepsilon\downarrow 0}\frac{1}{2\varepsilon}\mu[S_i(\omega)\cap(t-\varepsilon,t+\varepsilon)]=1\right\}\in\overline{\mathscr{B}\times\mathscr{F}}.$$

Let $\Lambda(t)=\{\omega: x(t,\omega)=i;\ (t,\omega)\notin\Gamma\}$ and $A(\omega)=\{t: x(t,\omega)=i;\ (t,\omega)\notin\Gamma\}$. Since $S_i(\omega)$ is a Lebesgue measurable set for a.a. ω, we have $\mu[A(\omega)]=0$ for a.a. ω by the density theorem in measure theory (see e.g. SAKS [1; p. 129]). Hence by FUBINI's theorem we have $\boldsymbol{P}\{\Lambda(t)\}=0$ for a.a. t. To prove (2) for all t we show that the left member of (2) does not depend on t by using the following lemma which is of general applicability.

Lemma. Let $\{\xi_t^{(\nu)}, t\in\boldsymbol{T}\}$, $\nu=1,2$, be two measurable processes whose sample functions are (almost all) Lebesgue integrable and suppose that the two processes have the same finite-dimensional joint distributions. Let (a_j,b_j), $1\leqq j\leqq k$ be any intervals. Then the two sets of k random variables $\int_{a_j}^{b_j}\xi^{(\nu)}(t,\omega)\,dt$, $1\leqq j\leqq k$, have the same joint distribution.

The lemma follows from a known result (see e.g. Doob [4; pp. 64—5]) according to which there exists a sequence of partitions $\{s_j^{(n)}\}$ with $a = s_0^{(n)} < s_1^{(n)} < \cdots < s_n^{(n)} = b$ such that

$$\int_a^b \xi^{(\nu)}(t, \omega)\, dt = \lim_{n\to\infty} \sum_{j=0}^{n-1} \xi^{(\nu)}(s_j^{(n)}, \omega)\,(s_{j+1}^{(n)} - s_j^{(n)})$$

for a.a. ω and $\nu = 1, 2$.

Let $\Delta(t) = \{\omega: x(t, \omega) = i\}$, $\mathsf{M}(t) = \{\omega: (t, \omega) \in \Gamma\}$ so that the probability in (2) reduces to $\boldsymbol{P}\{\mathsf{M}(t) | \Delta(t)\}$. As a consequence of the Lemma, the conditional joint distribution of $\mu[S_i(\omega) \cap (t-\varepsilon, t+\varepsilon)]$ for any finite number of values of ε, relative to $\Delta(t-\delta)$, where $0 < \varepsilon < \delta$, is the same for all $t > \delta$, owing to the stationarity of the transition probabilities. Hence in particular the probability

$$(3) \quad \boldsymbol{P}\{\Delta(t-\delta) \cap \Delta(t) \cap \mathsf{M}(t)\} / \boldsymbol{P}\{\Delta(t-\delta)\} = \boldsymbol{P}\{\Delta(t) \cap \mathsf{M}(t) | \Delta(t-\delta)\}$$

is the same for all $t > \delta$. Letting $\delta \downarrow 0$ it follows from the stochastic continuity of $\{x_t\}$ that the ratio in (3) tends to $\boldsymbol{P}\{\Delta(t) \cap \mathsf{M}(t)\} / P\{\Delta(t)\} = \boldsymbol{P}\{\mathsf{M}(t) | \Delta(t)\}$. Thus the latter is the same for all $t > 0$, as was to be shown. □

Theorem 3 may be stated as follows: if a given t belongs to $S_i(\omega)$, then it is a *point of density* of $S_i(\omega)$ for a.a. ω. A linear set S is called *metrically dense in itself* iff every open interval which contains a point of S contains a subset of S of positive measure.

Theorem 4 (SM). For a.a. ω, the set $S_i(\omega)$, $i \neq \infty$, is metrically dense in itself.

Proof. Let R be a separability set. For a.a. ω, an open interval which contains a point of $S_i(\omega)$ contains also a point of $R \cap S_i(\omega)$. This follows from the definition of separability if we take the closed set A in (4.10) to be $\bar{\boldsymbol{I}} - \{i\}$. On the other hand, applying Theorem 3 to $t = r$ for every $r \in R$ we see that for a.a. ω, an open interval containing a point of $R \cap S_i(\omega)$ contains a subset of $S_i(\omega)$ of positive measure (indeed on either side of r, if $r \neq 0$). Theorem 4 follows from these two statements. □

Theorem 5 (S). For any $i \in \boldsymbol{I}$, $s \geqq 0$ and $t > 0$, we have

$$(4) \qquad \boldsymbol{P}\{x(u, \omega) \equiv i,\ s < u < s+t \mid x(s, \omega) = i\} = e^{-q_i t}$$

where q_i is defined in (3.1), and $e^{-q_i t} = 0$ if $q_i = \infty$.

Proof. We have by (4.10)

$$\{\omega: x(r, \omega) \equiv i,\ r \in R \cap (s, s+t)\} - \{\omega: x(u, \omega) \equiv i,\ u \in (s, s+t)\} \subset \mathsf{N}.$$

The first set above belongs to $\mathscr{F}$ since R is denumerable, hence the second set also belongs to $\mathscr{F}$ since N is a null set and $(\mathscr{F}, \boldsymbol{P})$ is complete. Thus the probability on the left side of (4) is defined. To evaluate it we enumerate the set $R\cap(s, s+t)$ as $\{r_n\}$ and for each n, let $\{r_\nu, 1\leqq\nu\leqq n\}$ be rearranged into an increasing sequence $\{s_\nu^{(n)}, 1\leqq\nu\leqq n\}$. We have then

$$\boldsymbol{P}\{x(r,\omega)\equiv i, r\in R\cap(s,s+t)\}=\lim_{n\to\infty}\boldsymbol{P}\{x(s_\nu^{(n)},\omega)=i, 1\leqq\nu\leqq n\}.$$

Let $s_1^{(n)}-s=\delta_1^{(n)}$, $s_{\nu+1}^{(n)}-s_\nu^{(n)}=\delta_{\nu+1}^{(n)}$, $1\leqq\nu<n$; then the last written probability under the hypothesis $x(s,\omega)=i$ is equal to $\prod_{\nu=1}^{n} p_{ii}(\delta_\nu^{(n)})$. By Theorem 2.4, if $q_i<\infty$, we have

$$p_{ii}(\delta)=1-q_i\,\delta+o(\delta), \qquad \delta\downarrow 0;$$

consequently

$$\lim_{n\to\infty}\prod_{\nu=1}^{n} p_{ii}(\delta_\nu^{(n)})=\lim_{n\to\infty}\prod_{\nu=1}^{n}[1-q_i\,\delta_\nu^{(n)}+o(\delta_\nu^{(n)})]=\lim_{n\to\infty}e^{-q_i\sum_{\nu=1}^{n}\delta_\nu^{(n)}}=e^{-q_i t}$$

ince $\max_{1\leqq\nu\leqq n}\delta_\nu^{(n)}\to 0$ as $n\to\infty$. On the other hand if $q_i=\infty$, then for an arbitrary $A>0$, $p_{ii}(\delta)\leqq 1-A\,\delta$ if δ is sufficiently small, and the above limit is not greater than e^{-At} for every $A>0$, hence equal to zero. Therefore (4) is true in either case. ▯

Note that the above proof can be slightly simplified if we use well-separability which is implied by separability and stochastic continuity (Doob [4; Theorem II.2.2(i)]), the latter being equivalent to the standardness of the transition matrix by Theorem 4.3. We may then, for given s and t, take R to include all numbers of the form $s+m2^{-n}t$ and evaluate the left member of (4) as follows:

$$\lim_{n\to\infty}p_{ii}\left(\frac{t}{2^n}\right)^{2^n}=e^{-q_i t}.$$

Theorem 6 (S). If $q_i=\infty$, the probability is equal to zero that $S_i(\omega)$ contains any open interval. If $q_i<\infty$, the conditional probability under the hypothesis that $x(s,\omega)=i$ is equal to one that $S_i(\omega)$ contains an open interval containing s.

Proof. If $q_i=\infty$ and r is a rational number, it follows from Theorem 5 that the probability is equal to zero that $r\in S_i(\omega)$ and $(r, r+\varepsilon)\subset S_i(\omega)$ for any $\varepsilon>0$. Hence $S_i(\omega)$ does not contain any open interval since such an interval must contain a rational number. If $q_i<\infty$ it follows from Theorem 5 that for $0<\varepsilon<s$,

$$\boldsymbol{P}\{x(t,\omega)\equiv i,\ s-\varepsilon<t<s+\varepsilon\,|\,x(s,\omega)=i\}=\frac{p(s-\varepsilon)\,e^{-2q_i\varepsilon}}{p_i(s)}.$$

As $\varepsilon \downarrow 0$ this probability tends to one by the continuity of $p_i(\cdot)$. The limit is clearly the probability that $S_i(\omega)$ contains an open interval containing s. ▯

The state $i \in \boldsymbol{I}$ will be called *stable* if $q_i < \infty$, and *instantaneous* if $q_i = \infty$. If i is stable we define an *i-interval of the sample function* $x(\cdot, \omega)$ to be an open interval contained in $S_i(\omega)$ which is maximal in the sense that it is not a proper subset of another open interval contained in $S_i(\omega)$. The closure of an i-interval is called the corresponding *closed i-interval*; thus by definition a closed i-interval does not reduce to a point. When the state i is not specified these intervals will be referred to as *(closed) stable intervals*. The following theorem is of basic importance.

Theorem 7 (S). For a.a. ω, the set $\overline{S_i(\omega)}$ for each stable i is the union of all closed i-intervals, which are mutually disjoint and whose number is finite in every finite t-interval.

Proof. We prove first that there is a set $\Omega_1 \in \mathscr{F}$ with $\boldsymbol{P}(\Omega_1) = 1$ such that if $\omega \in \Omega_1$, then for a fixed stable i and every $A > 0$, the set $S_i(\omega)$ contains only a finite number of i-intervals in $(0, A)$. Consider for each n the sequence of numbers

$$\sigma_n = \left\{\frac{A}{n}, \frac{2A}{n}, \ldots, \frac{(n-1)A}{n}\right\}. \tag{5}$$

Since the union of all numbers of σ_n, $n \geqq 1$, is denumerable it follows from Theorem 6 that for any $t \in \bigcup_{n=1}^{\infty} \sigma_n$, the set $\{\omega\colon x(t, \omega) = i\}$ differs from the set $\{\omega\colon t$ belongs to an i-interval of $x(\cdot, \omega)\}$ by a null set. We define a sequence of ω-functions as follows:

$\tau_1^{(n)}(\omega) =$ the smallest number in σ_n that belongs to an i-interval of $x(\cdot, \omega)$, if such a number exists; otherwise $\tau_1^{(n)}(\omega) = A$;

$\tau_{s+1}^{(n)}(\omega) =$ the smallest number in σ_n that exceeds $\tau_s^{(n)}(\omega)$ and belongs to an i-interval of $x(\cdot, \omega)$ distinct from the one containing $\tau_s^{(n)}(\omega)$, if such a number exists; otherwise $\tau_{s+1}^{(n)}(\omega) = A$.

In particular if $\tau_s^{(n)}(\omega) = A$ then all the subsequent ones are equal to A. It is easy to verify that $\{\tau_s^{(n)}, s \geqq 1\}$ are random variables. Let N be a positive integer. Consider the ω-set

$$\Lambda_N^{(n)} = \{\omega\colon \tau_N^{(n)}(\omega) < A\};$$

thus $\Lambda_N^{(n)}$ is empty if $n \leqq N$. Let Λ_N be the ω-set such that there are at least N i-intervals of $x(\cdot, \omega)$ interior to $(0, A)$. For each ω in Λ_N there exists an $n = n(\omega)$ such that $\omega \in \Lambda_N^{(2^n)}$. Since $\Lambda_N^{(2^n)} \subset \Lambda_N^{(2^{n+1})}$ we have $\Lambda_N \subset \bigcup_{n=1}^{\infty} \Lambda_N^{(2^n)}$.

For $1 \leqq k \leqq n-1$ we put

$$\vartheta_k = \boldsymbol{P}\Big\{\frac{kA}{n} \text{ is the smallest number in } \sigma_n \text{ belonging to an } i\text{-interval}$$
$$\text{such that } x(t,\omega) \not\equiv i,\ 0 < t < \frac{kA}{n} \,\Big|\, x(0,\omega) = i\Big\}.$$

In words, ϑ_k is the probability that, starting from i at time 0, kA/n is the first instant t among the sequence σ_n such that $x(t,\omega)=i$ *after having left i some time between* 0 *and t*. This probability is well defined by separability and we have if $1 \leqq j < k \leqq n-1$,

$$\vartheta_{k-j} = \boldsymbol{P}\left\{\tau_{s+1}^{(n)}(\omega) = \frac{kA}{n} \,\Big|\, \tau_s^{(n)}(\omega) = \frac{jA}{n}\right\}. \tag{6}$$

To see this we observe that

$$\left\{\omega : \tau_s^{(n)}(\omega) = \frac{jA}{n}\right\} \in \left\{\omega : x\left(\frac{jA}{n}, \omega\right) = i\right\} \cap \mathscr{F}\left\{x_t,\, t < \frac{jA}{n}\right\};$$

hence by the Markov property the right member of (6) is equal to

$$\boldsymbol{P}\Big\{\frac{kA}{n} \text{ is the smallest number among the numbers of } \sigma_n \text{ exceeding } \frac{jA}{n}$$
$$\text{that belongs to an } i\text{-interval and such that } x(t,\omega) \not\equiv i,$$
$$\frac{jA}{n} < t < \frac{kA}{n} \,\Big|\, x\left(\frac{jA}{n}, \omega\right) = i\Big\}.$$

The last-written probability is equal to ϑ_{k-j} by stationary and separability (cf. the proof of Theorem 5), thus proving (6). Now the random variables $\{\tau_s^{(n)},\ 1 \leqq s \leqq N\}$ are easily seen to form a d.p.M.C. whose state space consists of the numbers of σ_n and the number A. If $\vartheta(\cdot,\cdot)$ denotes the transition matrix of this d.p.M.C., we have

$$\begin{aligned}
&\vartheta\left(\frac{jA}{n}, \frac{kA}{n}\right) = 0, && 1 \leqq k \leqq j \leqq n-1;\\
&\vartheta\left(\frac{jA}{u}, \frac{kA}{n}\right) = \vartheta_{k-j}, && 1 \leqq j < k \leqq n-1;\\
&\vartheta\left(\frac{jA}{n}, A\right) = 1 - \sum_{k=j+1}^{n-1} \vartheta_{k-j}, && 1 \leqq j \leqq n-1;\\
&\vartheta(A, A) = 1.
\end{aligned}$$

We have therefore

$$\begin{aligned}
\boldsymbol{P}(\Lambda_N^{(n)}) &= \sum_{1 \leqq k_1 < \cdots < k_N \leqq n-1} \boldsymbol{P}\left\{\tau_s^{(n)}(\omega) = \frac{k_s A}{n},\ 1 \leqq s \leqq N\right\}\\
&= \sum_{1 \leqq k_1 < \cdots < k_N \leqq n-1} \boldsymbol{P}\left\{\tau_1^{(n)}(\omega) = \frac{k_1 A}{n}\right\} \vartheta_{k_2-k_1} \cdots \vartheta_{k_N - k_{N-1}}\\
&\leqq \sum_{k=1}^{n-1} \boldsymbol{P}\left\{\tau_1^{(n)}(\omega) = \frac{kA}{n}\right\} \left(\sum_{k=1}^{n-1} \vartheta_k\right)^{N-1} \leqq \left(\sum_{k=1}^{n-1} \vartheta_k\right)^{N-1}.
\end{aligned}$$

Since

$$\sum_{k=1}^{n-1} \vartheta_k \leqq \boldsymbol{P}\{x(t,\omega) \not\equiv i,\ 0<t<A \mid x(0,\omega)=i\}$$

we have by Theorem 5,

$$\boldsymbol{P}(\Lambda_N^{(n)}) \leqq (1-e^{-q_i A})^{N-1}$$

regardless of the value of n. Hence letting $n\to\infty$ we obtain

$$\boldsymbol{P}(\Lambda_N) \leqq (1-e^{-q_i A})^{N-1}.$$

Let $\Lambda_\infty(A) = \bigcap_{N=1}^{\infty} \Lambda_N$; then $\Lambda_\infty(A)$ is the ω-set such that there are infinitely many i-intervals of $x(\cdot,\omega)$ in $(0, A)$. It follows from the above that $\boldsymbol{P}[\Lambda_\infty(A)]=0$ for every $A>0$. Let $\Lambda = \bigcup_{n=1}^{\infty} \Lambda_\infty(n)$; then $\boldsymbol{P}(\Lambda)=0$ and the set $\Omega_1 = \Omega - \Lambda$ fulfills our requirement.

By Theorem 6, if i is stable then for a.a. ω, each $r \in R \cap S_i(\omega)$ is contained in an i-interval of $x(\cdot,\omega)$. Hence by Theorem 4.3 there is a subset Ω_0 of Ω_1 with $\boldsymbol{P}(\Omega_0)=1$ such that if $\omega\in\Omega_0$, then (a) $x(\cdot,\omega)$ is separable in $\boldsymbol{T}$; and (b) the set $R\cap S_i(\omega)$ is contained in the union of i-intervals of $x(\cdot,\omega)$. Let $\omega\in\Omega_0$, $\tau\in S_i(\omega)$, $\tau\notin R$. Since $x(\cdot,\omega)$ is separable at τ, τ must be a limit point of the set $R\cap S_i(\omega)$. By the choice of Ω_0 there must then be i-intervals intersecting every neighborhood of τ. The number of such intervals being finite since $\omega\in\Omega_1$, a single interval must intersect every neighborhood of τ and hence τ is contained in the corresponding closed i-interval. We have therefore proved that every point of $S_i(\omega)$, whether in R or not, is contained in a closed i-interval, provided that $\omega\in\Omega_0$. Since there is only a finite number of i-intervals in every finite t-interval, the closure $\overline{S_i(\omega)}$ is the union of the closed i-intervals. By the choice of Ω_0, two i-intervals cannot abut at a point t since $x(t,\omega)$ must then have the value i by separability and so the two intervals must merge into one. Hence the closed i-intervals are all disjoint. Theorem 7 is completely proved. □

As a consequence of Theorem 7, for each stable i the successive i-intervals can be labelled as a finite or denumerable sequence; see § 15.

Notes. The idea of considering $u(x(t,\omega))$ in the proof of Theorem 2 is due to AUSTIN who used in addition the strong Markov property (see § 9 below); the present proof is a substantial simplification. A partial proof had previously been obtained by LÉVY [1] which will be given as Theorem 15.3 here. He later amended his proof of the general case in [5]. Historically, LÉVY attributed his fundamental paper [1] to a proof of this result. Theorem 3 was given in CHUNG [4] but the proof contained a mistake (the first inequality in (4.6) there). Theorem 4 is due to LÉVY [1] and Theorems 5 and 6 to DOOB [1]. Theorem 7 was stated by LÉVY [1] as a corollary to another result (Corollary 2 to Theorem 15.2 below) but should precede the latter; the proof is taken from CHUNG [4].

§ 6. Continuity properties of sample functions

In what follows we shall give further theorems concerning almost all sample functions. It is important to distinguish between two kinds of assertions: (i) an assertion about $x(t, \omega)$ for a.a. ω at a *fixed* t which is the same for all ω; (ii) an assertion about $x(\cdot, \omega)$ for a.a. ω at a *generic* t regarded as the running abscissa in the sample graph $(t, x(t, \omega))$, $t \in \boldsymbol{T}$, for each ω. Thus the exceptional null set may depend on t in case (i) but not in case (ii). Naturally (i) is a special case of (ii). As an example, the second assertion in Theorem 5.6 may be stated as follows: for a fixed t and stable i, $t \in S_i(\omega)$ implies that t is in an (open) i-interval of $x(\cdot, \omega)$ for a.a. ω. This is an assertion of the first kind; the corresponding assertion of the second kind is false at a generic t which is an endpoint of an i-interval. Similarly, Theorem 2 below is an assertion of the first kind while the corresponding assertion of the second kind will be given in Theorem 7.4.

The proof of the following basic theorem makes use of the theory of martingales.

Theorem 1 (S). For a.a. ω, the sample function $x(\cdot, \omega)$ has the following property. For any (generic) t, as $s \downarrow t$ [or $s \uparrow t$], $x(s, \omega)$ has at most one finite limiting value.

Proof. Let $A > 0$ and $j \in \boldsymbol{I}$ be given and define

$$\xi(\omega) = \begin{cases} 1 & \text{if } x(A, \omega) = j, \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, for $0 \leqq t < A$ define

$$y^A(t, \omega) = p_{x(t, \omega), j}(A - t). \tag{1}$$

Consider the stochastic process $\{y_t^A, 0 \leqq t < A\}$; we have by the Markov property of $\{x_t\}$, for a.a. ω,

$$\boldsymbol{E}\{\xi \mid x(s, \omega), 0 \leqq s \leqq t\} = \boldsymbol{E}\{\xi \mid x(t, \omega)\} = \boldsymbol{P}\{x(A, \omega) = j \mid x(t, \omega)\} = y^A(t, \omega).$$

Hence $\{y_t^A, 0 \leqq t < A\}$ is a martingale, being the conditional expectation of ξ relative to the nondecreasing sequence of Borel fields $\mathscr{F}\{x_s, 0 \leqq s \leqq t\}$, $t \geqq 0$. We now take a separable version of this martingale, without changing the notation, so that for each t, (1) remains true for a.a. ω. According to a martingale theorem (DOOB [4; Theorem VII.11.5]), for a.a. ω, the sample functions $y^A(\cdot, \omega)$ have left-hand and right-hand limits

$$(2) \qquad \lim_{s\uparrow t} y^A(s, \omega) = y^A(t-0, \omega), \qquad \lim_{s\downarrow t} y^A(s, \omega) = y^A(t+0, \omega)$$

for all t. Moreover, there is a denumerable set $D \subset (0, A)$ not depending on ω such that if $t \notin D$, then for a.a. ω

$$(3) \qquad y^A(t-0, \omega) = y^A(t, \omega) = y^A(t+0, \omega).$$

We are going to prove that in the present case the set D is empty. Because of the existence of the limits in (2) it is sufficient to show that for each fixed t, there exists a sequence $t_n \downarrow t$ [or $t_n \uparrow t$] such that $y^A(t_n, \omega) \to y^A(t, \omega)$ for a.a. ω. Now by stochastic continuity (Theorem 4.3), $x_s \to x_t$ in probability as $s \to t$; hence a subsequence $x_{t_n} \to x_t$ a.e. Since $x_t(\omega) \in \boldsymbol{I}$ and each element of $\boldsymbol{I}$ is isolated, we have $x(t_n, \omega) = x(t, \omega)$ for all sufficiently large n (depending on ω) and consequently by the continuity of p_{ij},

$$\lim_{n\to\infty} y^A(t_n, \omega) = \lim_{n\to\infty} p_{x(t_n, \omega), j}(A - t_n) = p_{x(t, \omega), j}(A - t) = y^A(t, \omega)$$

for a.a. ω. This concludes the proof that D is empty.

Put for $t \geqq 0$ ($0-0$ being always taken as 0):

$$(4) \qquad \eta^A(t, \omega) = y^A(t-0, \omega), \qquad \zeta^A(t, \omega) = y^A(t+0, \omega)$$

wherever these limits exist, namely a.e., and put η^A and ζ^A equal to zero elsewhere. Now consider the stochastic processes $\{\eta_t^A, 0 \leqq t < A\}$ and $\{\zeta_t^A, 0 \leqq t < A\}$. Each of these is a standard modification of $\{y_t^A, 0 \leqq t < A\}$ on account of (3) which has just been shown to hold for every fixed t. It is an elementary analytical consequence of the definitions in (4) that almost all sample functions $\eta^A(\cdot, \omega)$ are right continuous and almost all sample functions $\zeta^A(\cdot, \omega)$ are left continuous. Applying this result to all rational values of A, we conclude that there is a set Ω_0 with $\boldsymbol{P}(\Omega_0) = 1$ having the following properties. If $\omega \in \Omega_0$ then

(a) equation (1) holds for all $t \in R$ and all rational A;

(b) for all rational A, $\eta^A(\cdot, \omega)$ is right continuous and $\zeta^A(\cdot, \omega)$ is left continuous;

(c) for all rational A and all $r \in R$, $\eta^A(r, \omega) = y^A(r, \omega) = \zeta^A(r, \omega)$;

(d) $x(\cdot, \omega)$ is separable with respect to R in $\boldsymbol{T}$;

where R is a separability set. The property (d) is possible by Theorem 4.1. Now suppose for some $\omega_0 \in \Omega_0$ there exists a t such that $x(s, \omega_0)$ as $s \downarrow t$ has two finite limiting values i and i'. Then by (d) there exists $r_n \in R$,

$r'_n \in R$, $r_n \downarrow t$, $r'_n \downarrow t$ such that $x(r_n, \omega_0) = i$ and $x(r'_n, \omega_0) = i'$. We have by (c), (a) and the continuity of p_{ij} and $p_{i'j}$, for every rational $A > t$,

$$\lim_{n\to\infty} \eta^A(r_n, \omega_0) = \lim_{n\to\infty} y^A(r_n, \omega_0) = \lim_{n\to\infty} p_{ij}(A - r_n) = p_{ij}(A - t),$$

$$\lim_{n\to\infty} \eta^A(r'_n, \omega_0) = \lim_{n\to\infty} y^A(r'_n, \omega_0) = \lim_{n\to\infty} p_{i'j}(A - r'_n) = p_{i'j}(A - t).$$

Both limits are equal to $\eta^A(t, \omega_0)$ by (b); hence

$$p_{ij}(A - t) = p_{i'j}(A - t).$$

This being true for all rational $A > t$ we have, upon letting $A \downarrow t$, $\delta_{ij} = \delta_{i'j}$. This being true for every j, we have $i = i'$. The theorem is therefore proved in the case $s \downarrow t$; the proof in the case $s \uparrow t$ is similar if we use ζ_t^A instead of η_t^A. □

Theorem 2 (S). Let $t \geqq 0$ be fixed. Then the following assertions are true for a.a. ω:

(i) If $x(t, \omega) = i$ where i is stable, then $x(s, \omega) \to i$ as $s \to t$.

(ii) If $x(t, \omega) = i$ where i is instantaneous, then $x(s, \omega)$ has exactly two limiting values i and ∞ as $s \downarrow t$ [or $s \uparrow t$].

(iii) $x(t, \omega) \neq \infty$.

Proof. Assertion (iii) merely restates the fact that the adjoined state ∞ is fictitious. Thus (i) and (ii) are the only alternatives. Assertion (i) is in turn a restatement of the second part of Theorem 5.6. Finally if $x(t, \omega) = i$ then i is a limiting value of $x(s, \omega)$ as $s \downarrow t$ or $s \uparrow t$ by Theorem 5.3. There is no other finite limiting value by Theorem 1; and if i is instantaneous there must be another limiting value by the first part of Theorem 5.6. This limiting value can only be ∞, proving assertion (ii). □

Let us introduce the following new notation:

$$S_i^+(\omega) = \{t\colon\ S_i(\omega) \cap (t, t+\varepsilon) \neq 0 \quad \text{for every } \varepsilon > 0\};$$
$$S_i^-(\omega) = \{t\colon\ S_i(\omega) \cap (t-\varepsilon, t) \neq 0 \quad \text{for every } \varepsilon > 0\}.$$

Thus we have

$$\overline{S_i(\omega)} = S_i(\omega) \cup S_i^+(\omega) \cup S_i^-(\omega).$$

We write $\Lambda_1 \doteq \Lambda_2$ iff the two sets Λ_1 and Λ_2 differ by a null set; similarly $\Lambda_1 \mathrel{\dot{\subset}} \Lambda_2$ iff Λ_1 differs from a subset of Λ_2 by a null set. An immediate but important consequence of Theorem 2 can be stated as follows.

Theorem 3 (S). For each given t, we have if $i \neq \infty$:

$$\{\omega\colon t \in S_i(\omega)\} \doteq \{\omega\colon t \in \overline{S_i(\omega)}\} \doteq \{\omega\colon t \in S_i^+(\omega)\} \doteq \{\omega\colon t \in S_i^-(\omega)\}.$$

We are now in a position to deduce further properties of the sets of constancy beyond those given in §5.

Theorem 4 (S). For a.a. ω, if $i \neq j$ and both $\neq \infty$, then $\overline{S_i(\omega)} \cap \overline{S_j(\omega)}$ is a finite set in every finite t-interval.

Proof. Let the set of a.a. ω given in Theorem 1 be denoted by Ω_1, and let $B(\omega) = \overline{S_i(\omega)} \cap \overline{S_j(\omega)}$. If there is an $\omega_1 \in \Omega_1$ for which $B(\omega_1)$ is infinite in a finite t-interval, then there is a limit point t of $B(\omega_1)$ and an infinite subset of $B(\omega_1)$, all on the same side of t and converging to t. Thus as $s \to t$ from this side, $x(s, \omega_1)$ has the two limiting values i and j. This is impossible by the definition of Ω_1. ▯

Theorem 5 (SM). For a.a. ω, if $i \neq \infty$, $\overline{S_i(\omega)} - S_i(\omega)$ is of measure zero.

Proof. We have

$$\overline{S_i(\omega)} - S_i(\omega) = \bigcup_{j \neq i} [\overline{S_i(\omega)} \cap S_j(\omega)] \cup [\overline{S_i(\omega)} \cap S_\infty(\omega)].$$

The first union is at most denumerable by Theorem 4, and the last set is of measure zero by Theorem 5.1. The theorem follows. ▯

Theorem 5 will be sharpened later to Theorem 7.2 for a more specific version of the process.

Theorem 6 (SM). For a.a. ω, if i is instantaneous, the set $S_i(\omega)$ is nowhere dense in $\boldsymbol{T}$.

Proof. It follows from Theorem 5.6 that the set $S_i(\omega)$ does not contain any open interval. If $\overline{S_i(\omega)}$ contains an open interval then this interval must intersect some $S_j(\omega)$ with $j \neq i$; in fact it must intersect some $S_j(\omega)$ with $j \neq i$, $j \neq \infty$ by separability (Theorem 4.1) since ∞ is not an isolated value. By Theorem 5.4 it must then contain a subset of $S_j(\omega)$ of positive measure, contradicting Theorem 5. Hence $\overline{S_i(\omega)}$ does not contain any open interval, and so $S_i(\omega)$ is nowhere dense. ▯

Notes. Theorem 1 is due to DOOB [1]; the proof here is simpler because the martingale theorem used in that paper has been strengthened to yield convergence from both sides of t. Theorems 4, 5 and 6 were given by LÉVY [1], [3]; the proofs are taken from CHUNG [4].

§ 7. Further specifications of the process

The results of the last two sections are valid if the M.C. is separable and measurable. These two properties do not determine the sample functions with probability one. More precisely, given the initial distribution and the standard transition matrix, there are in general different versions of the process which are measurable and separable with the same separability set but in which sample functions corresponding to the same ω are different for a.a. ω. For example, the method of construction of a separable standard modification (DOOB [4; pp. 58—59]) allows the value of $x(\cdot, \omega)$ at a point of $\overline{S_i(\omega)} \cap \overline{S_j(\omega)}$, $i \neq j$, to be either i or j (but no other value) indifferently. Since the set $\{(t, \omega) : t \in \overline{S_i(\omega)} \cap \overline{S_j(\omega)}\}$ has $\mu \times \boldsymbol{P}$ measure zero by Theorem 6.4 and FUBINI's theorem, the value of $x(\cdot, \cdot)$ can be changed on any subset of it without affecting measurability. Similarly, at a point of $\overline{S_i(\omega)} \cap S_\infty(\omega)$ the value of $x(\cdot, \omega)$ may be i or ∞ indifferently. We are therefore still at liberty to specify versions of the M.C. in which the sample functions are determined with probability one and possess desirable properties.

From here on in the book let us suppose that $\boldsymbol{I}$ *is the set of positive integers, unless otherwise specified.*

As a consequence of Theorem 6.2, we have for each fixed t, and a.a. ω,

$$(1) \qquad \underline{\lim}_{s \to t} x(s, \omega) = \underline{\lim}_{s \downarrow t} x(s, \omega) = \underline{\lim}_{s \uparrow t} x(s, \omega) = x(t, \omega).$$

It is important to remark that the lower limits above are all taken in a *deleted* neighborhood of t. We now put

$$\underline{x}(t, \omega) = \underline{\lim}_{s \to t} x(s, \omega), \qquad x_+(t, \omega) = \underline{\lim}_{s \downarrow t} x(s, \omega), \qquad x_-(t, \omega) = \underline{\lim}_{s \uparrow t} x(s, \omega)$$

and $x_-(0, \omega) = x(0, \omega)$. It follows from (1) that each of the three processes

$$(2) \qquad \{\underline{x}(t), t \in \boldsymbol{T}\}, \qquad \{x_+(t), t \in \boldsymbol{T}\}, \qquad \{x_-(t), t \in \boldsymbol{T}\}$$

is a standard modification of $\{x_t, t \in \boldsymbol{T}\}$. We proceed to show that, if we redefine the sample functions on a null set, then each of the three processes in (2) has specific properties. The function x defined on $\boldsymbol{T}$ will be said to be *lower* [or *right lower*, or *left lower*] semicontinuous there iff $x = \underline{x}$ [or x_+, or x_-] on $\boldsymbol{T}$. Observe that this is a more stringent definition than the usual one where "$=$" is replaced by "$\leqq$".

Theorem 1. Let an arbitrary stochastic process $\{x_t, t \in \boldsymbol{T}\}$ be separable with separability set R and satisfy the condition that for each fixed t,

$$(3) \qquad \boldsymbol{P}\left\{\underline{\lim}_{s \downarrow t} x(s, \omega) = x(t, \omega)\right\} = 1.$$

Then there is a version $\{\tilde{x}_t, t \in \boldsymbol{T}\}$ of the process which is separable with respect to R and Borel measurable, and in which all sample functions are right lower semicontinuous. We have $\tilde{x}(\cdot, \omega) \equiv x_+(\cdot, \omega)$ for a.a. ω. These assertions remain true, *mutatis mutandis*, for the other two lower limits.

Proof. It follows from separability and (3) that there exists a set Ω_0 with $\boldsymbol{P}(\Omega_0)=1$ having the following two properties. If $\omega\in\Omega_0$,

(a) $x(\cdot,\omega)$ is separable with respect to R in $\boldsymbol{T}$;

(b) $x(r,\omega)=x_+(r,\omega)$ for every $r\in R$.

We define $\tilde{x}(t,\omega)\equiv x_+(t,\omega)$ if $\omega\in\Omega_0$; and $\tilde{x}(t,\omega)\equiv i$ if $\omega\in\Omega-\Omega_0$ where i is a fixed state. Then $\{\tilde{x}_t\}$ is a version of $\{x_t\}$ since $\{x_+(t)\}$ is. If $\omega\in\Omega_0$, then for any t, the value of $x_+(t,\omega)$ lies in the closure of values of $x(r,\omega)$, $r\neq t$, $r\in R\cap(t,t+n^{-1})$ by (a); hence $x_+(\cdot,\omega)$ is separable in $\boldsymbol{T}$ with respect to R by (b). If $\omega\in\Omega-\Omega_0$, $\tilde{x}(\cdot,\omega)$ is well-separable in $\boldsymbol{T}$ by definition. Hence R is a separability set for $\{\tilde{x}_t\}$. Next we have by (a), if $\omega\in\Omega_0$, and $R=\{r_j\}$,

$$x_+(t,\omega)=\lim_{n}\inf_{t<r_j<t+n^{-1}} x(r_j,\omega).$$

Let $e(r,t)=1$ if $t<r<t+n^{-1}$, and $=0$ otherwise; let $z(r,\omega)=1$ if $x(r,\omega)<c$, and $=0$ otherwise, where c is any real number. Clearly we have

$$\left\{(t,\omega):\inf_{t<r_j<t+n^{-1}} x(r_j,\omega)<c\right\}=\left\{(t,\omega):\sum_j e(r_j,t)\,z(r_j,\omega)>0\right\}.$$

Each $e(r_j,\cdot)$ is measurable $\mathscr{B}$ and each $z(r_j,\cdot)$ is measurable $\mathscr{F}$. Hence the above set belongs to $\mathscr{B}\times\mathscr{F}$. Since c is arbitrary it follows that $x_+(\cdot,\cdot)$ on $\boldsymbol{T}\times\Omega_0$ is measurable $\mathscr{B}\times\mathscr{F}$. Hence $\tilde{x}(\cdot,\cdot)$ on $\boldsymbol{T}\times\Omega$ is measurable $\mathscr{B}\times\mathscr{F}$. Finally if $\omega\in\Omega_0$, we have for every t,

$$x_+(t,\omega)=\varliminf_{r_j\downarrow t} x(r_j,\omega)=\varliminf_{r_j\downarrow t} x_+(r_j,\omega)=\varliminf_{s\downarrow t} x_+(s,\omega)$$

by (a), (b) and the separability of $\{x_+(t)\}$ already proved. Thus $x_+(\cdot,\omega)$, for each $\omega\in\Omega_0$, is right lower semicontinuous and so is $\tilde{x}(\cdot,\omega)$ for each $\omega\in\Omega$. This proves the assertions for $\{x_+(t)\}$; the proof for $\{x_-(t)\}$ is exactly the same and that for $\{\underline{x},(t)\}$ then follows since $\underline{x}=\min(x_+,x_-)$. ▯

Remark. It follows from the proof that as a function on $[0,t)\times\Omega$, the function $x(\cdot,\cdot)$ is measurable $\mathscr{B}_t\times\mathscr{F}_t^0$ where $\mathscr{B}_t=[0,t)\cap\mathscr{B}$ and $\mathscr{F}_t^0=\mathscr{F}\{x_s,0\leqq s\leqq t\}$. Such a process is called *progressively naturally Borel measurable*, see CHUNG and DOOB [1].

The restriction of Ω to Ω_0 being absolutely immaterial for all purposes we shall refer to the three versions of the M.C. given in Theorem 1 simply by the notation in (2) or as $\underline{x}$, x_+ and x_-. They are all well-separable since x may be taken to be so. For each of these, almost every sample function is completely determined by its values on any separability set, and the sets of constancy $S_i(\omega)$ have special properties. For instance, $t\in S_\infty(\omega)$ if and only if $\underline{x}(s,\omega)\to\infty$ as $s\to t$; $x_+(s,\omega)\to\infty$ as $s\downarrow t$; $x_-(s,\omega)\to\infty$ as $s\uparrow t$; respectively. Theorem 6.5 can now be strengthened as follows.

Theorem 2. For the $\underline{x}$ version the set $\overline{S_i(\omega)} \cap S_\infty(\omega)$, $i \neq \infty$, is empty; for the x_+ or x_- version it is denumerable for a.a. ω. For any of these three versions, the set $\overline{S_i(\omega)} - S_i(\omega)$, $i \neq \infty$, is at most denumerable for a.a. ω.

Proof. The first assertion follows at once from the remark preceding the theorem. By the same remark, each point of $\overline{S_i(\omega)} \cap S_\infty(\omega)$ is isolated on the right or left side according as x_+ or x_- is in question. Such a set is clearly at most denumerable. Now the proof of Theorem 6.5 shows that $\overline{S_i(\omega)} - S_i(\omega)$ is also at most denumerable. ▯

Theorem 3. For any of the three versions $\underline{x}$, x_+ or x_-, we have if $i \neq \infty$:

$$S_i^+(\omega) \cap S_i^-(\omega) \subset S_i(\omega) \subset S_i^+(\omega) \cup S_i^-(\omega);$$

the second inclusion being true for any separable and measurable version.

Proof. For a.a. ω, if $t \in S_i^+(\omega) \cap S_i^-(\omega)$, then i is the only finite limiting value of $x(\cdot, \omega)$ at t by Theorem 6.1. Hence $t \in S_i(\omega)$ by the definition of the three versions. Conversely if $t \in S_i(\omega)$, then since $x(\cdot, \omega)$ is separable at t for a.a. ω by Theorems 4.1 and 4.3, we have $t \in S_i^+(\omega) \cup S_i^-(\omega)$. ▯

A more complete analysis of the limiting behavior of the sample function at a generic t can now be given, elaborating Theorem 6.1.

Theorem 4 (SM). Almost all sample functions have the following property. For any (generic) t, as $s \downarrow t$ [or $s \uparrow t$] there are three possibilities:

(a) $x(s, \omega) \to i$, where i is stable;

(b) $x(s, \omega)$ has exactly two limiting values i and ∞, where i is instantaneous;

(c) $x(s, \omega) \to \infty$.

Furthermore:

(α) if $x(t, \omega) = i$ where i is stable, then (a) is true (with the same i) as $s \to t$ from at least one side;

(β) if $x(t, \omega) = i$ where i is instantaneous, then (b) is true (with the same i) as $s \to t$ from at least one side;

Finally for the three versions $\underline{x}$, x_+ and x_-:

(γ) if $x(t, \omega) = \infty$ then (c) is true from at least one side.

Proof. That (a), (b) and (c) are the only possibilities without specifying the nature of i is the assertion of Theorem 6.1. In case there is a unique finite limit i this cannot be instantaneous by Theorem 5.6, hence it must be stable. In case there are two limiting values i and ∞ we have $t \in \overline{S_i(\omega)}$ and yet $\overline{S_i(\omega)}$ does not contain a closed interval with t as one endpoint;

hence i cannot be stable by Theorem 5.7 and must therefore be instantaneous. The assertions (α) and (β) follow from separability and (γ) follows from the definition of the three versions. $\square$

Another way to summarize the situation is to examine the set of discontinuity $D(\omega)$ of the sample function $x(\cdot, \omega)$. For the sake of expediency let us call a point t such that $\lim\limits_{s \to t} x(s, \omega) = \infty$ a point of discontinuity even if $x(t, \omega) = \infty$ as required by separability. The set of such t will be denoted by $S_\infty^*(\omega)$.

Theorem 5. For any of the three versions $\underline{x}$, x_+ or x_-, and a.a. ω, the set of continuity of the sample function $x(\cdot, \omega)$ is the union of all (open) stable intervals, if any. The set of discontinuity is the union of the following sets:

(a) the sets $\overline{S_i(\omega)}$ where i is instantaneous, each being a perfect set and having positive measure;

(b) the set $S_\infty^*(\omega)$ of measure zero;

(c) the sets $\overline{S_i(\omega)} \cap S_\infty(\omega)$ where i is stable, each being a finite set in every finite t-interval;

(d) the sets $\overline{S_i(\omega)} \cap \overline{S_j(\omega)}$ where $i \neq j$ are both stable, each being a finite set in every finite t-interval.

A point belonging to a set in the category (d) is called a *point of jump*. The sets in (d) are mutually disjoint and also disjoint from the sets in the other categories. More generally, if i, j and ∞ are distinct then

$$\overline{S_i(\omega)} \cap \overline{S_j(\omega)} \cap S_\infty(\omega) = 0$$

for any of the three versions. (This need not be true for another version.) Thus the sets in (c) are mutually disjoint; a set in (a) and one in (c) are disjoint. For the version $\underline{x}$, we have $S_\infty^*(\omega) = S_\infty(\omega)$ and the sets in (c) are empty.

An important special case of a M.C. with a standard transition matrix is where all the states are stable, namely $q_i < \infty$ for all $i \in \boldsymbol{I}$. In this case the sample function behavior is simpler since case (b) of Theorem 4 is ruled out.

Corollary (SM). If all states are stable then $D(\omega)$ is a closed set of measure zero for a.a. ω. At a point of $D(\omega)$ which is not a point of jump $x(\cdot, \omega) \to \infty$ from at least one side.

The first assertion in the corollary can be proved directly as follows. The set

$$\left\{(t, \omega) : \lim_{s \to t} x(s, \omega) = x(t, \omega)\right\}$$

belongs to $\overline{\overline{\mathscr{B}\times\mathscr{F}}}$ since the process is separable and measurable. Its section for each fixed t has probability one by Theorem 6.2(i). Hence by Fubini's theorem its section for almost every fixed ω differs from $\boldsymbol{T}$ by a set of measure zero.

It is interesting to note that the set $D(\omega)$ in the Corollary above need not contain any point of jump; see Example 7 of § 20.

Notes. Theorem 1 is taken from Chung [6], with an important revision. Borel measurability is frequently postulated in the discussion of certain processes in the literature, though not in Doob [4]. Of course if all sample functions are sufficiently smooth, say right continuous, then Borel measurability follows from measurability since $x(t,\omega)=\lim\limits_{n\to\infty} x\big((n+1)[tn^{-1}],\omega\big)$. A new result (Chung and Doob [1]) published after the first edition states that every stochastically continuous process has a well-separable and progressively naturally Borel measurable version. This can be used to replace Theorem 1 here and simplify proofs elsewhere.

Theorem 4 is a more specific version of a result due to Doob [1], made possible by the more detailed analysis of stable vs. instantaneous states given in the preceding theorems. The Corollary to Theorem 5 is explicitly proved by Lévy [1] in which some other results are implicit.

§ 8. Optional random variable

We begin by recalling the definition of a conditional expectation relative to a given field. Given the probability triple $(\Omega, \mathscr{F}, \boldsymbol{P})$, a random variable ζ with $\boldsymbol{E}(|\zeta|)<\infty$ and an augmented Borel subfield $\mathscr{G}$ of $\mathscr{F}$, any ω-function $\chi(\cdot)$ which is measurable $\mathscr{G}$ and such that

$$\int_{\mathsf{M}} \chi(\omega)\boldsymbol{P}(d\omega)=\int_{\mathsf{M}} \zeta(\omega)\boldsymbol{P}(d\omega)$$

for every $\mathsf{M}\in\mathscr{G}$, is called a *version of the conditional expectation of* ζ *relative to* $\mathscr{G}$, and denoted collectively by $\boldsymbol{E}(\zeta|\mathscr{G})$. Thus $\tilde{\chi}$ is another version of the conditional probability if and only if $\tilde{\chi}$ is measurable $\mathscr{G}$ and $\chi=\tilde{\chi}$ with probability one.

When $\mathscr{G}$ is the field generated by the random variable α the conditional expectation $\boldsymbol{E}(\zeta|\mathscr{G})$ will be denoted by $\boldsymbol{E}(\zeta|\alpha)$ and its value at ω by $\boldsymbol{E}(\zeta|\alpha(\omega))$. It is well known (see e.g. Doob [4; p. 603]) that one version of $\boldsymbol{E}(\zeta|\alpha)$ is a Baire function of α; when so regarded its value on the set $\{\omega: \alpha(\omega)=s\}$ will be denoted by $\boldsymbol{E}(\zeta|\alpha=s)$. When ζ is the indicator of a set Λ in $\mathscr{F}$, the *conditional probability* of Λ relative to $\mathscr{G}$ is denoted by $\boldsymbol{P}(\Lambda|\mathscr{G})$ or $\boldsymbol{P}(\Lambda|\alpha)$, etc.

Let $\{x_t, t \in \boldsymbol{T}\}$ be a M.C. and let $\{\mathscr{F}_t, t \in \boldsymbol{T}\}$ be a family of Borel fields such that

$$(1) \qquad \mathscr{F}_s \subset \mathscr{F}_t \text{ if } s \leqq t; \quad \mathscr{F}\{x_s, 0 \leqq s \leqq t\} \subset \mathscr{F}_t \subset \mathscr{F};$$

and such that for every $j \in \boldsymbol{I}$ and $0 < s < t$ we have

$$(2) \qquad \boldsymbol{P}\{x_t(\omega) = j \mid \mathscr{F}_s\} = p_{x_s, j}(t-s) = \boldsymbol{P}\{x_t(\omega) = j \mid x_s\}$$

with probability one. From this it follows more generally that if $\Lambda \in \mathscr{F}\{x_t, t \geqq s\}$, then

$$\boldsymbol{P}\{\Lambda \mid \mathscr{F}_s\} = \boldsymbol{P}\{\Lambda \mid x_s\}$$

with probability one. Observe that if $\mathscr{F}_t = \mathscr{F}\{x_s, 0 \leqq s \leqq t\}$ then (1) is satisfied and (2) is a consequence of the Markov property. Now let $0 \leqq r < t$ and let $s \downarrow r$ through rational values in (2), then the middle member converges with probability one to $p_{x_r, j}(t-r)$ by (6.2) and the argument following it, while the left member becomes $\boldsymbol{P}\{x_t(\omega) = j \mid \mathscr{F}_{r+0}\}$ where $\mathscr{F}_{r+0} = \bigcap_{s > r} \mathscr{F}_s$. This being true for every r we conclude that (2) remains valid if $\mathscr{F}_s$ is replaced by the generally larger field $\hat{\mathscr{F}}_s = \mathscr{F}_{s+0}$. The new family of Borel fields $\{\hat{\mathscr{F}}_t, t \in \boldsymbol{T}\}$ satisfies also (1) and is such that $\hat{\mathscr{F}}_t = \hat{\mathscr{F}}_{t+0}$. Replacing the given family by the new one without changing the notation, we shall suppose in what follows that the family $\{\mathscr{F}_t, t \in \boldsymbol{T}\}$ satisfies (1) and (2) together with the additional condition of *right continuity:*

$$\mathscr{F}_t = \mathscr{F}_{t+0} = \bigcap_{s > t} \mathscr{F}_s, \quad t \in \boldsymbol{T}.$$

Such a family of Borel fields will be called *admissible* relative to $\{x_t, t \in \boldsymbol{T}\}$. The most important one is $\{\mathscr{F}_t^0, t \in \boldsymbol{T}\}$, where $\mathscr{F}_t^0 = \mathscr{F}\{x_s, 0 \leqq s \leqq t\}$; in this case there is right continuity as it stands by the theorem below.

Theorem 1. $\mathscr{F}_{t+0}^0 = \mathscr{F}_t^0$ for each $t \in \boldsymbol{T}$.

Proof. According to the preceding remarks we have, if $\Lambda \in \mathscr{F}\{x_r, r > t\}$ then

$$(3) \qquad \boldsymbol{P}\{\Lambda \mid \mathscr{F}_t^0\} = \boldsymbol{P}\{\Lambda \mid \mathscr{F}_{t+0}^0\},$$

with probability one. Now let Λ be an "elementary" set in $\mathscr{F}_\infty^0$ of the form $\Lambda = \{\omega\colon x_{t_\nu}(\omega) = i_\nu, 1 \leqq \nu \leqq l\}$. Then we have $\Lambda = \Lambda_1 \cap \Lambda_2$ where $\Lambda_1 \in \mathscr{F}_t^0$ and $\Lambda_2 \in \mathscr{F}\{x_r, r > t\}$. It follows that if φ denotes the indicator of Λ_1:

$$\boldsymbol{P}\{\Lambda \mid \mathscr{F}_t^0\} = \varphi \boldsymbol{P}\{\Lambda_2 \mid \mathscr{F}_t^0\} = \varphi \boldsymbol{P}\{\Lambda_2 \mid \mathscr{F}_{t+0}^0\} = \boldsymbol{P}\{\Lambda \mid \mathscr{F}_{t+0}^0\}$$

with probability one. The class of sets Λ satisfying (3) is a monotone class containing the field generated by the elementary sets, hence it contains the least Borel field $\mathscr{F}_\infty^0$ so generated. In particular if $\Lambda \in \mathscr{F}_{t+0}^0$, then the right member in (3) reduces to the indicator of Λ while the left member is measurable with respect to $\mathscr{F}_t^0$. Hence $\Lambda \in \mathscr{F}_t^0$ since all fields are augmented by assumption, proving the theorem. ▯

The family $\{\mathscr{F}_t^0, t \in \boldsymbol{T}\}$ will be called the *minimal admissible family* relative to the M.C. $\{x_t, t \in \boldsymbol{T}\}$. Let $\mathscr{F}_t$ be the smallest Borel field generated by $\mathscr{F}_t^0$ and random variables independent of $\{x_t, t \in \boldsymbol{T}\}$, then $\{\mathscr{F}_t, t \in \boldsymbol{T}\}$ is an admissible family which is larger than the minimal one.

Let $\Delta \in \mathscr{F}$ and $\boldsymbol{P}(\Delta) > 0$. A nonnegative extended-valued random variable α defined on Δ is said to be *optional relative to* $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$, where $\{\mathscr{F}_t, t \in \boldsymbol{T}\}$ is an admissible family, iff for each $t \geqq 0$ we have $\{\omega: \alpha(\omega) < t\} \in \mathscr{F}_t$. Because $\mathscr{F}_{t+0} = \mathscr{F}_t$ the last condition is equivalent to

$$\{\omega: \alpha(\omega) \leqq t\} \in \mathscr{F}_t. \tag{4}$$

According to this definition, α is optional relative to $\{x_t, \mathscr{F}_t\}$, where $\{\mathscr{F}_t\}$ is any admissible family, if it is optional relative to $\{x_t, \mathscr{F}_t^0\}$. In the latter case we shall say that α is *optional relative to* $\{x_t, t \in \boldsymbol{T}\}$ or simply "optional". On the set $\Omega - \Delta$ the random variable α may be undefined so that the pertinent probability triple for α is the reduced triple $(\Delta, \Delta\mathscr{F}, \boldsymbol{P}(\cdot \mid \Delta))$. If α is optional relative to $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$ it is easily seen that the collection of sets Λ in $\Delta\mathscr{F}$ such that

$$\Lambda \cap \{\omega: \alpha(\omega) \leqq t\} \in \mathscr{F}_t \tag{5}$$

for every $t \geqq 0$ forms a Borel field. This will be called the *pre-α field* (relative to $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$) and denoted by $\mathscr{F}_\alpha$. Note that when α reduces to a constant $t_0 \geqq 0$ (which is an optional random variable relative to any admissible family) the corresponding $\mathscr{F}_\alpha$ reduces indeed to $\mathscr{F}_{t_0}$. In the particular case of the minimal admissible family the pre-α field may be denoted by $\mathscr{F}\{x_t, t \leqq \alpha\}$.

In the remainder of this section the M.C. $\{x_t, t \in \boldsymbol{T}\}$ is assumed to be separable and measurable. Since we have already assumed that the transition matrix is standard, or equivalently (Theorem 4.3) that the process is stochastically continuous, this implies by a theorem quoted before (see the paragraph preceding Theorem 4.3) that the process is well-separable. In particular, the set of terminating dyadics $\{m 2^{-n}, m, n \geqq 0\}$ or the set of rational numbers is a separability set. The first set is especially convenient since as n increases, the partial sets $\{m 2^{-n}, m \geqq 0\}$ give rise to partitions of $\boldsymbol{T}$ which are refinements of preceding ones.

Let α with domain of finiteness Δ be optional relative to $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$. For each $j \in \boldsymbol{I}$ we define a ω-function γ_j as follows:

$$\gamma_j(\omega) = \inf\{t: t > \alpha(\omega); x(t, \omega) = j\}. \tag{6}$$

In this and subsequent similar definitions it is understood that the infimum is ∞ if no t exists satisfying the given condition. Since for a.a. ω, $x(\cdot, \omega)$ is separable in $\boldsymbol{T}$ by Theorem 4.1, we infer $\gamma_j(\omega)$ is a limit point of $R \cap S_j(\omega)$ from the right, where R is a separability set. The domain of definition and finiteness of γ_j is

$$\Gamma_j = \Delta \cap \{\omega\colon S_j(\omega) \cap (\alpha(\omega), \infty) \neq 0\}.$$

We have, if $\{s_\nu\}$ is a separability set,

$$\{\omega\colon \gamma_j(\omega) < t\} = \bigcup_{s_\nu < t} \{\omega\colon \alpha(\omega) < s_\nu;\ x(s_\nu, \omega) = j\}.$$

Hence γ_j is optional.

Let Λ be an arbitrary set in the pre-α field $\mathscr{F}_\alpha$ and put

$$A(\Lambda; s) = \boldsymbol{P}\{\Lambda; \alpha(\omega) \leqq s\}.$$

For a fixed Λ the function $A(\Lambda; \cdot)$ is thus a distribution function, giving rise to a measure on $\mathscr{B}$ to be called the $A(\Lambda; \cdot)$ measure. As a general notation the integral of a function C over B with respect to a measure $A(\cdot)$ on $\mathscr{B}$ will be denoted by

$$\int_B C(s) A(ds).$$

Where the range of α is in question "almost all s" or "a.a. s" will mean all $s \in \boldsymbol{T}$ except a set of $A(\Lambda; \cdot)$ measure zero.

The conditional probability distribution in the wide sense of γ_j with respect to Λ and α is a function $C_j(\cdot, \cdot | \Lambda)$ of the point $s \in \boldsymbol{T}$ and the set $B \in \mathscr{B}$ such that

(a) for each s, $C_j(s, \cdot | \Lambda)$ is a measure on $\mathscr{B}$;
(b) for each B, $C_j(\cdot, B | \Lambda)$ is a Baire function on $\boldsymbol{T}$;
(c) for every $B_1, B_2 \in \mathscr{B}$ we have

$$\int_{B_1} C_j(s, B_2) | \Lambda) A(\Lambda, ds) = \boldsymbol{P}\{\Lambda; \alpha(\omega) \in B_1; \gamma_j(\omega) \in B_2\}.$$

Such a C_j exists (see Doob [4; p. 29]). In particular, for a fixed $B \in \mathscr{B}$, $C_j(\alpha(\omega), B | \Lambda)$ is the conditional probability of the set $\{\omega\colon \gamma_j(\omega) \in B\}$ relative to the smallest augmented Borel field containing Λ and the sets $\{\omega\colon \alpha(\omega) \leqq s\}$, $s \geqq 0$; thus

$$C_j(s, B | \Lambda) = \boldsymbol{P}\{\gamma_j(\omega) \in B | \Lambda; \alpha = s\} \tag{7}$$

for a.a. s. When the set B is the interval $[0, u]$ we shall write $C_j(s, u | \Lambda)$ for $C_j(s, B | \Lambda)$.

If K and Λ belong to $\mathscr{F}$ with $\boldsymbol{P}(\Lambda) > 0$ and χ is a random variable, the conditional probability of K relative to χ on the couple $(\Lambda\mathscr{F}, \boldsymbol{P}(\cdot | \Lambda))$

will be denoted by $\boldsymbol{P}\{\mathsf{K}|\Lambda;\chi\}$; similarly for more than one random variable or for conditional expectation. Thus we have, cancelling a factor $\boldsymbol{P}(\Lambda)^{-1}$ from both sides,

$$\boldsymbol{P}\{\mathsf{K};\Lambda;\chi(\omega)\leqq t\}=\int\limits_{\Lambda\cap\{\omega:\,\chi(\omega)\leqq t\}}\boldsymbol{P}\{\mathsf{K}|\Lambda;\chi\}\boldsymbol{P}(d\omega).$$

Theorem 2. *Let $j,k\in\boldsymbol{I}$, $t\geqq 0$; and $\Lambda\in\mathscr{F}_\alpha$, $\boldsymbol{P}(\Lambda)>0$. We have almost everywhere on the set $\{\omega:\gamma_j(\omega)\leqq t\}$,*

$$\text{(8)}\qquad \boldsymbol{P}\{x(t,\omega)=k|\Lambda;\alpha,\gamma_j\}=p_{jk}(t-\gamma_j).$$

Proof. For any s and s' in $[0,t]$ we have

$$\begin{aligned}\Lambda\cap\{\omega:\alpha(\omega)\leqq s;\gamma_j(\omega)\leqq s';\,x(t,\omega)=k\}&\\ =(\Lambda_1\cup\Lambda\cup_2\Lambda_3)\cap\{\omega:\,x(t,\omega)=k\}&\end{aligned}$$

where

$$\begin{aligned}\Lambda_1&=\Lambda\cap\{\omega:\alpha(\omega)\leqq s;\gamma_j(\omega)\leqq s';\gamma_j(\omega)<t\},\\ \Lambda_2&=\Lambda\cap\{\omega:\alpha(\omega)\leqq s;\gamma_j(\omega)\leqq s';\alpha(\omega)=\gamma_j(\omega)=t\},\\ \Lambda_3&=\Lambda\cap\{\omega:\alpha(\omega)\leqq s;\gamma_j(\omega)\leqq s';\alpha(\omega)<\gamma_j(\omega)=t\}.\end{aligned}$$

We have

$$\{\omega:\alpha(\omega)<\gamma_j(\omega)=t\}\in\{\omega:\,t\in\overline{S_j(\omega)}\}-\{\omega:\,t\in S_j^-(\omega)\},$$

hence this is a null set by Theorem 6.3 and *a fortiori* $\boldsymbol{P}(\Lambda_3)=0$. By the same theorem, we have $\{\omega:\gamma_j(\omega)=t\}\in\{\omega:\,x(t,\omega)=j\}$ and consequently

$$\boldsymbol{P}\{\Lambda_2;\,x(t,\omega)=k\}=\delta_{jk}\boldsymbol{P}(\Lambda_2)=\int\limits_{\Lambda_2}p_{jk}(t-\gamma_j(\omega))\,\boldsymbol{P}(d\omega).$$

It remains to evaluate $\boldsymbol{P}\{\Lambda_1;\,x(t,\omega)=k\}$. For this purpose we define for each $n\geqq 0$,

$$\gamma_j^{(n)}(\omega)=\min\left\{\frac{m}{2^n}:\frac{m}{2^n}>\gamma_j(\omega);\;x\left(\frac{m}{2^n},\omega\right)=j\right\}.$$

Clearly $\gamma_j^{(n)}$ is a random variable; by the definition of γ_j and well-separability we have $\gamma_j^{(n)}(\omega)\downarrow\gamma_j(\omega)$ for a.a. $\omega\in\Gamma_j$. Hence we have

$$\begin{aligned}&\boldsymbol{P}\{\Lambda_1;\,x(t,\omega)=k\}\\ &\quad=\lim_{n\to\infty}\boldsymbol{P}\{\Lambda;\alpha(\omega)\leqq s;\gamma_j(\omega)\leqq s';\gamma_j^{(n)}(\omega)<t;\,x(t,\omega)=k\}\\ &\quad=\lim_{n\to\infty}\sum_{m<t2^n}\boldsymbol{P}\left\{\Lambda;\alpha(\omega)\leqq s;\gamma_j(\omega)\leqq s';\gamma_j^{(n)}(\omega)=\frac{m}{2^n};\,x(t,\omega)=k\right\}.\end{aligned}$$

Since α and γ_j are both optional and $\alpha\leqq\gamma_j\leqq\gamma_j^{(n)}$, it is easy to see that

$$\Lambda\cap\left\{\omega:\alpha(\omega)\leqq s;\gamma_j(\omega)\leqq s';\gamma_j^{(n)}(\omega)=\frac{m}{2^n}\right\}\in\mathscr{F}_{m2^{-n}}.$$

Therefore we have by (2),

$$\begin{aligned}
&\boldsymbol{P}\{\Lambda_1;\, x(t,\omega)=k\}\\
&\quad=\lim_{n\to\infty}\sum_{m<t2^n}\boldsymbol{P}\left\{\Lambda;\,\alpha(\omega)\leqq s;\,\gamma_j(\omega)\leqq s';\,\gamma_j^{(n)}(\omega)=\frac{m}{2^n}\right\}p_{jk}\left(t-\frac{m}{2^n}\right)\\
&\quad=\lim_{n\to\infty}\int\limits_{\Lambda\cap\{\omega:\,\alpha(\omega)\leqq s;\,\gamma_j(\omega)\leqq s';\,\gamma_j^{(n)}(\omega)<t\}} p_{jk}\left(t-\gamma_j^{(n)}(\omega)\right)\boldsymbol{P}(d\omega)\\
&\quad=\int\limits_{\Lambda_1} p_{jk}\left(t-\gamma_j(\omega)\right)\boldsymbol{P}(d\omega).
\end{aligned}$$

Combining the above results we obtain

$$\begin{aligned}
\boldsymbol{P}\{\Lambda;\,\alpha(\omega)\leqq s;\,\gamma_j(\omega)\leqq s';\,x(t,\omega)=k\}&=\int\limits_{\Lambda_1\cup\Lambda_2\cup\Lambda_3} p_{jk}\left(t-\gamma_j(\omega)\right)\boldsymbol{P}(d\omega)\\
&=\int\limits_{\Lambda\cap\{\omega:\,\alpha(\omega)\leqq s;\,\gamma_j(\omega)\leqq s'\}} p_{jk}\left(t-\gamma_j(\omega)\right)\boldsymbol{P}(d\omega).
\end{aligned}$$

The truth of this equation for arbitrary s and s' in $[0, t]$ is precisely the assertion (8). ▯

Theorem 3. One version of the conditional probability

(9) $$\boldsymbol{P}\{x(t,\omega)=j\,|\,\Lambda;\,\alpha=s\}$$

is given by

(10) $$r_j(s,t\,|\,\Lambda)=\int\limits_{[s,t]} p_{jj}(t-u)\,C_j(s,du\,|\,\Lambda),\qquad j\in\boldsymbol{I},\ 0\leqq s\leqq t.$$

For each $s\geqq 0$, $r_j(s,\cdot\,|\,\Lambda)$ is right continuous in $[s,\infty)$; and $r_j(\cdot,\cdot\,|\,\Lambda)$ is a Baire function of the pair (s,t) in $0\leqq s\leqq t$.

Proof. We have with probability one,

(11) $$\boldsymbol{P}\{x(t,\omega)=j\,|\,\Lambda;\,\alpha\}=\boldsymbol{E}\{\boldsymbol{P}[x(t,\omega)=j\,|\,\Lambda;\,\alpha,\gamma_j]\,|\,\Lambda;\,\alpha\}.$$

By (7) and (8) with $j=k$, the right member of (11) is equal to that of (10) for $\alpha=s$ and a.a. s; hence the latter is a version of (9). For each s, $C_j(s,u\,|\,\Lambda)$ is right continuous in u by definition, hence $r_j(s,t\,|\,\Lambda)$ is right continuous in t by (10), p_{jj} being continuous. By the same token $r_j(s,t\,|\,\Lambda)$ can be expressed as the limit of Riemann-Stieltjes sums of Baire functions of (s,t), hence it is a Baire function of (s,t). ▯

In what follows a.a. t or (t,t') refer to the Lebesgue measure on $\boldsymbol{T}$ or $\boldsymbol{T}\times\boldsymbol{T}$. If a relation involving several variables holds under several conditions, they are to be satisfied in the order of their appearance, and the "a.a." set in a later condition may depend on the variables in the preceding conditions. For example, by "$t, t'\geqq 0$, a.a. $s\in[0,t]$" we mean "for each $t\geqq 0$ and $t'\geqq 0$, and each $s\in[0,t]-Z(t,t')$ where $Z(t,t')$ has A-measure zero".

Theorem 4. We have for every $k \in \boldsymbol{I}$,

$$r_k(s, t+t' \mid \Lambda) = \sum_j r_j(s, t \mid \Lambda)\, p_{jk}(t'), \qquad \text{a.a. } s,\ t > s,\ t' \geqq 0; \tag{12}$$

and

$$\sum_j r_j(s, t \mid \Lambda) = 1, \qquad \text{a.a. } s,\ t > s. \tag{13}$$

For every $j \in \boldsymbol{I}$ and a.a. s, the function $r_j(s, \cdot \mid \Delta)$ is continuous in $[s, \infty]$.

Proof. Since α is optional, we have if $0 \leqq s \leqq t$, $t' > 0$,

$$\boldsymbol{P}\{\Lambda;\ \alpha(\omega) \leqq s;\ x(t+t', \omega) = k\} = \sum_j \boldsymbol{P}\{\Lambda;\ \alpha(\omega) \leqq s;\ x(t, \omega) = j\}\, p_{jk}(t').$$

This being true for every s, it follows from the defining property of a conditional probability that

$$\boldsymbol{P}\{x(t+t', \omega) = k \mid \Lambda;\ \alpha\} = \sum_j \boldsymbol{P}\{x(t, \omega) = j \mid \Lambda;\ \alpha\}\, p_{jk}(t')$$

with probability eon; that is, by Theorem 3,

$$r_k(s, t+t' \mid \Lambda) = \sum_j r_j(s, t \mid \Lambda)\, p_{jk}(t') \tag{14}$$

for $t, t' \geqq 0$, a.a. $s \in [0, t]$. Both members of (14) being Baire functions of (s, t, t') by Theorem 3, we can apply FUBINI's theorem and infer that (14) is also true for a.a. s, a.a. $(t, t') \in [s, \infty) \times [0, \infty)$. From this it follows, by the right continuity given in Theorem 3 and FATOU's lemma:

$$r_k(s, t+t' \mid \Lambda) \geqq \sum_j r_j(s, t \mid \Lambda)\, p_{jk}(t'), \qquad \text{a.a. } s,\ t \geqq s,\ t' \geqq 0. \tag{15}$$

Summing over k, we obtain

$$\sum_k r_k(s, t+t' \mid \Lambda) \geqq \sum_j r_j(s, t \mid \Lambda). \tag{16}$$

Next, we have from Theorem 3,

$$\sum_j r_j(s, t \mid \Lambda) = \boldsymbol{P}\{\Lambda \mid \Lambda;\ \alpha = s\} = 1 \tag{17}$$

for $t > 0$, a.a. $s \in [0, t]$. Hnece by FUBINI's theorem (17) is also true for a.a. s, a.a. $t \geqq s$. In the same way as (16) follows from (14), we have from (17),

$$\sum_j r_j(s, t \mid \Lambda) \leqq 1, \qquad \text{a.a. } s,\ t \geqq s. \tag{18}$$

Consider a fixed $s \notin E$ where E is the exceptional set of A-measure zero occurring in (15) and (18). For such an s if (17) is true for a certain value of t then it is true for all larger values of t by (16) and (18). Hence (17) is in fact true for all $t > s$, since it is true for a.a. $t \geqq s$. This proves

(13). Hence there must be equality in (16) and so also in (15). This proves (12). The continuity of $r_j(s, \cdot|\Lambda)$ in (s, ∞), for a.a. s, now follows from (12), (13) and the continuity of p_{jk}. Since for each s it is right continuous at s by Theorem 3, the continuity may be extended to $[s, \infty)$. ▯

Notes. An optional random variable is called a "stopping time" in Blumenthal [1] and a "random variable independent of the future" in Dynkin and Yuškevič [1]. In the latter as well as in Chung [6], however, the more exigent condition (4) is imposed. The relaxation to the present condition with strict inequality, which is more easily verified in applications, is made possible by Theorem 1 due to Doob and his observation (both communicated in a letter) that the right continuity of the field family may be assumed. These definitions extend at once to a Markov process with general state space; see the first two papers cited above, and also Yuškevič [1]. For a general discussion of questions concerning optionality see Chung and Doob [1].

Instead of defining a M.C. $\{x_t\}$ first and then an admissible family $\{\mathscr{F}_t\}$, we may define $\{x_t\}$ to be a M.C. *relative to* $\{\mathscr{F}_t\}$ iff the double family $\{x_t, \mathscr{F}_t\}$ satisfies (1) and (2) with a given transition matrix (p_{ij}). The larger the fields $\mathscr{F}_t$ are the more stringent this definition becomes, but at the same time the less stringent the definition of optionality relative to $\{x_t, \mathscr{F}_t\}$. This kind of generality is needed, though not in this monograph, when several processes are considered simultaneously.

The proof of Theorem 1 is new and much simpler than the one given in the first edition. Theorems 2 to 4 are taken from Chung [6]; Theorems 3 and 4 were first obtained in the special case treated in § 15.

§ 9. Strong Markov property

To begin with, we assume that the M.C. $\{x_t, t \in \boldsymbol{T}\}$ is Borel measurable. Using the notation of § 8, we may define a family of random variables $\{\xi_t, t \in \boldsymbol{T}\}$ on the triple $(\Delta, \Delta\mathscr{F}, \boldsymbol{P}(\cdot|\Delta))$ as follows:

$$(1) \qquad \xi_t(\omega) = \xi(t, \omega) = x(\alpha(\omega) + t, \omega).$$

For each t, $\xi_t(\cdot)$ is measurable $\mathscr{F}$ since $x(\cdot, \cdot)$ is measurable $\mathscr{B} \times \mathscr{F}$. In other words each ξ_t is a random variable, thugh not necessarily finite-valued. It will be proved later that except for ξ_0 which may be ∞ a.e. on Δ, each ξ_t is actually finite a.e. on Δ. The stochastic process $\{\xi_t, t \in \boldsymbol{T}\}$, taken in this general sense, is called the *post-α process relative to* $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$. The Borel field $\mathscr{F}\{\xi_t, t \in \boldsymbol{T}\}$ will be called the *post-α field* and denoted by $\mathscr{F}'_\alpha$. We remark that (1) defines a process if the x-process is an arbitrary Borel measurable process and $\alpha \geqq 0$; on the other hand, even if the x-process is not Borel measurable it may happen that (1) still defines a process (see e.g. CHUNG [5]). For intuitive background we refer to § I.13 where the discrete parameter analogue is discussed. Indeed, even in the continuous parameter case, if the optional random variable α is assumed to be denumerably-valued the results of § I.13 can be immedaitely extended. We leave this as an exercise since the following discussion covers the general situation.

Using the notation in § 8, we put for any $\Lambda \in \mathscr{F}_\alpha$:

$$r_j(\Lambda; t) = \int_0^\infty r_j(s, s+t|\Lambda) A(\Lambda; ds), \quad j \in \boldsymbol{I}, t \geqq 0. \tag{2}$$

It follows from Theorem 8.4 that for any j, $k \in \boldsymbol{I}$ and $t, t' > 0$, we have

$$r_j(\Lambda; t) \geqq 0; \quad \sum_j r_j(\Lambda; t) = \boldsymbol{P}(\Lambda); \quad \sum_j r_j(\Lambda; t) p_{jk}(t') = r_k(\Lambda; t+t'). \tag{3}$$

Moreover, $r_j(\Lambda; t)$ is continuous on $\boldsymbol{T}$; this can also be proved by applying Theorem 2.3 to the last relation in (3). In particular, $\lim_{t \downarrow 0} r_j(\Lambda; t) = r_j(\Lambda; 0)$. Hence by FATOU's lemma, the two equations in (3) become inequalities with "$\leqq$" when $t=0$.

The probabilistic meaning of $r_j(\Lambda; 0)$ will be given first.

Theorem 1. For the x_+ version, we have if $\Lambda \in \mathscr{F}_\alpha$,

$$\begin{cases} \boldsymbol{P}\{\Lambda; \xi_0(\omega) = j\} = r_j(\Lambda; 0), & j \in \boldsymbol{I}; \\ \boldsymbol{P}(\Lambda; \xi_0(\omega) = \infty\} = \boldsymbol{P}(\Lambda) - \sum_j r_j(\Lambda; 0). \end{cases} \tag{4}$$

Proof. Suppose that $\alpha(\omega) = \gamma_j(\omega)$ for a.a. ω in Δ, where γ_j is defined in (8.6). Then by Theorem 6.1, the only finite limiting value of $x(t, \omega)$ as $t \downarrow \alpha(\omega)$ is j and $\xi(0, \omega) = j$ for a.a. ω in Δ by the choice of the x_+ version. Conversely if $\xi(0, \omega) = j$ then $\alpha(\omega) = \gamma_j(\omega)$ for a.a. ω in Δ by the definition of γ_j. We have thus $\{\omega: \xi(0, \omega) = j\} \doteq \{\omega: \alpha(\omega) = \gamma_j(\omega)\}$. Consequently we obtain, using (8.7),

$$\begin{aligned} \boldsymbol{P}\{\Lambda; \xi(0, \omega) = j\} &= \boldsymbol{P}\{\Lambda; \alpha(\omega) = \gamma_j(\omega)\} \\ &= \lim_{n \to \infty} \sum_{m=0}^\infty \int_{\left[\frac{m}{n}, \frac{m+1}{n}\right)} C_j\left(s, \frac{m+1}{n} \middle| \Lambda\right) A(\Lambda; ds) \\ &= \lim_{n \to \infty} \int_0^\infty C_j\left(s, \frac{[ns+1]}{n} \middle| \Lambda\right) A(\Lambda; ds) = \int_0^\infty C_j(s, s|\Lambda) A(\Lambda; ds). \end{aligned}$$

Since $C_j(s, s|\Lambda) = r_j(s, s|\Lambda)$ by (8.10) the first equation in (4) is proved; the second then follows since $\xi(0, \omega) = \infty$ if and only if $\xi(0, \omega) \notin \boldsymbol{I}$. ☐

Theorem 1 is immediately generalized as follows.

Theorem 2. For the x_+ version, we have if $\Lambda\in\mathscr{F}_\alpha$; $0\leqq t_0<t_1<\cdots<t_N$; and $j_0, j_1, \ldots, j_N\in\boldsymbol{I}$:

$$(5)\qquad \boldsymbol{P}\{\Lambda;\, \xi(t_\nu,\omega)=j_\nu,\ 0\leqq\nu\leqq N\}=r_{j_0}(\Lambda;t_0)\prod_{\nu=0}^{N-1}p_{j_\nu j_{\nu+1}}(t_{\nu+1}-t_\nu).$$

Proof. Consider the set

$$B_n=\Lambda\cap\bigcup_{m=0}^{\infty}\left\{\omega:\frac{m}{n}\leqq\alpha(\omega)<\frac{m+1}{n};\ x\left(\frac{m+1}{n}+t_\nu,\,\omega\right)=j_\nu,\ 0\leqq\nu\leqq N\right\}.$$

If $\omega\in\limsup_n B_n$, then there is an infinite sequence of rational numbers $\{r_k\}$ such that $r_k\downarrow\alpha(\omega)$ and $x(r_k+t_\nu,\omega)=j_\nu$ for $0\leqq\nu\leqq N$. Hence by Theorem 6.1, for almost every such ω, we have $\xi(t_\nu,\omega)=\underline{\lim}_{t\downarrow\alpha(\omega)+t_\nu}x(t,\omega)=j_\nu$. It follows that

$$(6)\qquad \limsup_n B_n\subset\Lambda\cap\{\omega:\xi(t_\nu,\omega)=j_\nu,\ 0\leqq\nu\leqq N\}.$$

To evaluate $\boldsymbol{P}(B_n)$, we use Theorem 8.3:

$$\begin{aligned}\boldsymbol{P}(B_n)&=\sum_{m=0}^{\infty}\boldsymbol{P}\left\{\frac{m}{n}\leqq\alpha(\omega)<\frac{m+1}{n};\ x\left(\frac{m+1}{n}+t_0,\,\omega\right)=j_0\right\}\times\\&\qquad\times\boldsymbol{P}\left\{x\left(\frac{m+1}{n}+t_\nu,\,\omega\right)=j_\nu,\ 1\leqq\nu\leqq N\,\Big|\,x\left(\frac{m+1}{n}+t_0,\,\omega\right)=j_0\right\}\\&=\sum_{m=0}^{\infty}\ \int\limits_{\left[\frac{m}{n},\frac{m+1}{n}\right)}r_{j_0}\left(s,\,\frac{m+1}{n}+t_0\Big|\Lambda\right)A(\Lambda;ds)\,Q\\&=\int_0^{\infty}r_{j_0}\left(s,\,\frac{[ns+1]}{n}+t_0\Big|\Lambda\right)A(\Lambda;ds)\,Q\end{aligned}$$

where

$$(7)\qquad Q=\prod_{\nu=0}^{N-1}p_{j_\nu\,j_{\nu+1}}(t_{\nu+1}-t_\nu).$$

Hence we have, using Theorem 8.4 and bounded convergence,

$$(8)\qquad\begin{aligned}\lim_{n\to\infty}\boldsymbol{P}(B_n)&=\int_0^{\infty}\lim_{n\to\infty}r_{j_0}\left(s,\,\frac{[ns+1]}{n}+t_0\Big|\Lambda\right)A(\Lambda;ds)\,Q\\&=\int_0^{\infty}r_{j_0}(s,\,s+t_0|\Lambda)\,A(\Lambda;ds)\,Q=r_{j_0}(\Lambda;t_0)\,Q.\end{aligned}$$

It follows from (6) and (7) that

$$(9)\qquad \boldsymbol{P}\{\Lambda;\,\xi(t_\nu,\omega)=j_\nu,\ 0\leqq\nu\leqq N\}\geqq\lim_{n\to\infty}\boldsymbol{P}(B_n)=r_{j_0}(\Lambda;t_0)\,Q;$$

namely, (5) is true with "$=$" replaced by "$\geqq$". Summing this inequality over all $j_\nu \in \boldsymbol{I}$, $1 \leqq \nu \leqq N$, we obtain

$$(10)\quad \begin{cases} \boldsymbol{P}\{\Lambda;\, \xi(t_0,\omega)=j_0\} \\ \geqq \sum\limits_{j_1\in\boldsymbol{I}} \cdots \sum\limits_{j_N\in\boldsymbol{I}} \boldsymbol{P}\{\Lambda;\, \xi(t_0,\omega)=j_0;\, \xi(t_\nu,\omega)=j_\nu,\, 1\leqq \nu \leqq N\} \geqq r_{j_0}(\Lambda; t_0). \end{cases}$$

If $t_0=0$ the two extreme members in (10) are equal by (4); hence equality must hold in (5). If $t_0>0$, summing (10) over all $j_0 \in \boldsymbol{I}$ we have by the second relation in (3),

$$(11)\qquad \boldsymbol{P}(\Lambda) \geqq \sum_{j_0\in\boldsymbol{I}} \boldsymbol{P}\{\Lambda;\, \xi(t_0,\omega)=j_0\} \geqq \sum_{j_0\in\boldsymbol{I}} r_{j_0}(\Lambda; t_0) = \boldsymbol{P}(\Lambda).$$

It follows as before that equality must hold in (5). □

If α is optional so is $\alpha+t_0$ in the same sense for any constant $t_0 \geqq 0$; the pre-$(\alpha+t_0)$ field is $\mathscr{F}_{\alpha+t_0}$ and the post-$(\alpha+t_0)$ field is $\mathscr{F}'_{\alpha+t_0}$.

Corollary 1. For each $t\geqq 0$, ξ_t is measurable $\mathscr{F}_{\alpha+t}$.

Proof. Equality must hold throughout (9); this and (6) imply that the "$\dot{\in}$" in (6) can be replaced by "$\doteq$". Taking $\Lambda=\Delta$ and $N=0$ in the definition of B_n we see that $B_n \in \mathscr{F}_{\alpha+t_0+1/n}$. Hence $\limsup\limits_n B_n \in \mathscr{F}_{\alpha+t_0}$ by the right continuity of the field family. Thus we have $\Delta \cap \{\omega\colon \xi_{t_0}(\omega) = j_0\} \in \mathscr{F}_{\alpha+t_0}$ for every j_0, proving the corollary. □

Corollary 2. For each $t>0$, ξ_t is finite a.e. on Δ.

Proof. This follows from (11). □

Corollary 3. For $t_0>0$, the equation (5) remains true for the x_- and $\underline{x}$ versions.

Proof. We need only consider, instead of B_n, either

$$\Lambda \cap \bigcup_{m=0}^{\infty} \left\{\omega\colon \frac{m}{n} \leqq \alpha(\omega) < \frac{m+1}{n};\; x\left(\frac{m}{n}+t_\nu,\,\omega\right) = j_\nu,\; 0\leqq \nu \leqq N\right\}$$

or

$$\Lambda \cap \bigcup_{m=0}^{\infty} \left\{\omega\colon \frac{m}{n} \leqq \alpha(\omega) < \frac{m+1}{n};\right.$$
$$\left. x\left(\frac{m}{n}+t_\nu,\,\omega\right) = x\left(\frac{m+1}{n}+t_\nu,\,\omega\right) = j_\nu,\; 0\leqq \nu \leqq N\right\}$$

and use the last step in the proof of the theorem. □

The proof of Corollary 3 explains the choice of the x_+ version in order to have (5) at $t_0=0$.

Theorem 2 will now be formulated in a comprehensive manner with all details.

Theorem 3. Let the M.C. $\{x_t, t\in \boldsymbol{T}\}$ be the x_+ version with the compactified state space $\bar{\boldsymbol{I}}$ and the standard transition matrix (p_{ij}). For any α with domain of finiteness Δ and optional relative to $\{x_t, \mathscr{F}_t, t\in\boldsymbol{T}\}$, the post-$\alpha$ process $\{\xi_t, t\in\boldsymbol{T}\}$ defined in (1) is a M.C. in $\boldsymbol{T}$ on the probability triple $(\Delta', \Delta'\mathscr{F}, \boldsymbol{P}(\cdot|\Delta'))$ where $\Delta'=\{\omega: \xi_0(\omega)\neq\infty\}$; with a certain compactified state space $\bar{\boldsymbol{J}}\subset\bar{\boldsymbol{I}}$ where the adjoined state ∞, if present, remains the adjoined state; with the transition matrix (p_{ij}) restricted to $\boldsymbol{J}$; and the absolute distribution at $t\in\boldsymbol{T}$ given by $\{r_j(\Delta'; t), j\in\boldsymbol{I}\}$. Furthermore if $t_0\in\boldsymbol{T}$ and $\Lambda\in\mathscr{F}_{\alpha+t_0}$, $\mathsf{M}\in\mathscr{F}'_{\alpha+t_0}$, then

$$\boldsymbol{P}\{\Lambda\mathsf{M}|\xi_{t_0}\}=\boldsymbol{P}\{\Lambda|\xi_{t_0}\}\boldsymbol{P}\{\mathsf{M}|\xi_{t_0}\} \tag{12}$$

a.e. on Δ'. The family of Borel fields $\{\mathscr{F}_{\alpha+t}, t\in\boldsymbol{T}\}$ is admissible relative to $\{\xi_t, t\in\boldsymbol{T}\}$.

The same conclusions remain true if Δ' is replaced by Δ, provided that $\boldsymbol{T}$ is replaced by $\boldsymbol{T}^0$.

Remark. The minimal state space $\boldsymbol{J}$ of the post-α process is the set of $j\in\boldsymbol{I}$ for which there exists a $t\geqq 0$ such that $r_j(\Delta, t)>0$. It includes all j for which there exists an $i\in\boldsymbol{I}$ and a $t\geqq 0$ such that $r_i(\Delta', 0)>0$ and $p_{ij}(t)>0$.

Proof. That the adjoined state remains so is a consequence of Corollary 2 to Theorem 2. To prove (12) we note first that (5) implies the special case where $\Lambda\in\mathscr{F}_\alpha$ and $\mathsf{M}=\Delta'\cap\{\omega: \xi(t_\nu, \omega)=j_\nu, 0\leqq\nu\leqq N\}$, and consequently also for an arbitrary $\mathsf{M}\in\mathscr{F}\{\xi_t, t\geqq t_0\}$ by the properties of a Borel field. Now for $t_0>0$ let us put

$$\eta_t(\omega)=\xi_{t_0+t}(\omega)=x(\alpha(\omega)+t_0+t, \omega);$$

then $\{\eta_t, t\in\boldsymbol{T}\}$ is the post-$(\alpha+t_0)$ process and $\mathscr{F}\{\eta_t, t\geqq 0\}=\mathscr{F}'_{\alpha+t_0}$. Since $\eta_0=\xi_{t_0}$ is finite a.e. on Δ by Corollary 2 to Theorem 2, we may apply (5) to the η-process; thus if $\Lambda\in\mathscr{F}_{\alpha+t_0}$ and $\mathsf{M}\in\mathscr{F}'_{\alpha+t_0}$ then

$$\boldsymbol{P}\{\Lambda\mathsf{M}|\eta_0\}=\boldsymbol{P}\{\Lambda|\eta_0\}\boldsymbol{P}\{\mathsf{M}|\eta_0\}$$

a.e. on Δ. This proves (12) in full. Dividing (12) through by $\boldsymbol{P}(\Lambda|\xi_{t_0})$ provided the latter is not zero, we obtain

$$\boldsymbol{P}\{\mathsf{M}|\Lambda; \xi_{t_0}\}=\boldsymbol{P}\{\mathsf{M}|\xi_{t_0}\}.$$

The truth of this for every $t_0\in\boldsymbol{T}$, a.e. in Δ', and Corollary 1 to Theorem 2, imply the admissibility of $\{\mathscr{F}_{\alpha+t}, t\in\boldsymbol{T}\}$ relative to $\{\xi_t, t\in\boldsymbol{T}\}$.

The last assertion in Theorem 3 is implicit in the discussion above. ▯

It is easy to see that α is measurable $\mathscr{F}_\alpha$ so that for any s the set $\{\omega: \alpha(\omega)\leqq s\}$ belongs to $\mathscr{F}_\alpha$ and may be absorbed in the set Λ in (5). However, it is advantageous to make this more explicit.

Corollary 1. We have if $\Lambda \in \mathscr{F}_\alpha$, $0 \leqq t_0 < \cdots < t_N$ and $j_0, \ldots, j_N \in \boldsymbol{I}$,

$$(13) \quad \begin{cases} \boldsymbol{P}\{x(\alpha(\omega)+t_\nu, \omega) = j_\nu, 0 \leqq \nu \leqq N \mid \Lambda; \alpha\} \\ \qquad\qquad = r_{j_0}(\alpha, \alpha+t_0 \mid \Lambda) \prod_{\nu=0}^{N-1} p_{j_\nu j_{\nu+1}}(t_{\nu+1}-t_\nu) \end{cases}$$

a.e. on Δ' if $t_0 = 0$ and a.e. on Δ if $t_0 > 0$.

Proof. Replacing the Λ in (5) by $\Lambda_1 = \Lambda \cap \{\omega : \alpha(\omega) \leqq s\}$ as we may, and recalling the definition of $A(\Lambda_1; \cdot)$, we have

$$\begin{aligned} \boldsymbol{P}\{\Lambda; \alpha(\omega) \leqq s; \xi(t_\nu, \omega) = j_\nu, 0 \leqq \nu \leqq N\} &= r_{j_0}(\Lambda_1; t_0)\, Q \\ &= \int_0^\infty r_{j_0}(u, u+t_0 \mid \Lambda_1)\, Q A(\Lambda_1; du) \\ &= \int_{\Lambda \cap \{\omega : \alpha(\omega) \leqq s\}} r_{j_0}(\alpha(\omega), \alpha(\omega)+t_0 \mid \Lambda)\, Q \boldsymbol{P}(d\omega). \end{aligned}$$

The truth of this for every s is equivalent to (13). ▯

We recall (Theorem 8.3) that for a *fixed* t, $r_j(s, t \mid \Lambda) = \boldsymbol{P}\{x(t, \omega) = j \mid \Lambda; \alpha = s\}$ for a.a. s in $A(\Lambda; \cdot)$ measure. It would be meaningless to substitute $s+t$ for t in this equation, because the resulting right member would not be defined. A correct interpretation of $r_j(s, s+t \mid \Lambda)$ is given by (13), according to which it is equal to $\boldsymbol{P}\{x(\alpha(\omega)+t, \omega) = j \mid \Lambda; \alpha\}$ evaluated at $\alpha = s$ for a.a. s in $A(\Lambda; \cdot)$ measure. Viewed in this way, the corollary above expresses a *substitution property* of the conditional probability.

Corollary 2. The post-α process, in $\boldsymbol{T}$ on Δ' and in $\boldsymbol{T}^0$ on Δ, is Borel measurable and stochastically continuous. There is a version $\{\xi_+(t)\}$ [or $\{\xi_-(t)\}$, or $\{\underline{\xi}(t)\}$] of it which is well-separable and Borel measurable, with (almost) all sample functions right lower [or left lower, or lower] semicontinuous.

Proof. Borel measurability follows from that of the x-process and stochastic continuity follows from the standard transition matrix. The remaining assertion is a consequence of Theorem 7.1 applied to the M.C. $\{\xi_t\}$. ▯

It is important to note that the post-α process has not yet been shown to be separable *as it stands*. This will now be proved.

Corollary 3. The post-α process is well-separable (as it stands).

Proof. Let $\{\tilde{\xi}_t, t \in \boldsymbol{T}\}$ be a well-separable and measurable version of $\{\xi_t, t \in \boldsymbol{T}\}$ and let R be a denumerable set dense in $\boldsymbol{T}$. Then for a.a. ω (in Δ) we have $\xi(r, \omega) = \tilde{\xi}(r, \omega)$ for all $r \in R$. Next, since the set $\{(t, \omega) : \xi(t, \omega) \neq \tilde{\xi}(t, \omega)\}$ is measurable $\overline{\mathscr{B} \times \mathscr{F}}$ and its ω-section for each fixed t has probability zero, FUBINI's theorem implies that for a.a. ω the t-section $\{t: \xi(t, \omega) \neq \tilde{\xi}(t, \omega)\}$ has Lebesgue measure zero. Finally, since the graph of $\xi(\cdot, \omega)$ is a "tail part" of that of $x(\cdot, \omega)$, Theorem 5.4 implies that for a.a. ω the set $\breve{S}_i(\omega) = \{t: \xi(t, \omega) = i\}$ for each $i \in \boldsymbol{I}$ is metrically dense in itself. Combining these three statements we conclude that for a.a. ω, an open interval intersecting $\breve{S}_i(\omega)$ intersects also $\tilde{S}_i(\omega) = \{t: \tilde{\xi}(t, \omega) = i\}$ and therefore, since $\{\tilde{\xi}_t\}$ is separable (with respect to R), it intersects $R \cap \tilde{S}_i(\omega) = R \cap \breve{S}_i(\omega)$. Thus (see § 4) for each ω, $\xi(\cdot, \omega)$ is separable in the set $\bigcup_{i \in \boldsymbol{I}} \breve{S}_i(\omega)$. Now by Theorem 7.4($\gamma$) it follows easily that $\xi(\cdot, \omega)$ is also separable in $\breve{S}_\infty(\omega)$, hence altogether in $\boldsymbol{T}$, proving the corollary. ▯

An instructive way to view Theorem 3 is as follows. Let $j \in \bar{\boldsymbol{I}}$, $\Delta_j = \Delta \cap \{\xi_0(\omega) = j\}$ and define $\alpha_j(\omega) = \alpha(\omega)$ on Δ_j. It is easy to see that each α_j is optional relative to $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$. The post-$\alpha_j$ process is simply the post-α process restricted to the triple $(\Delta_j, \Delta_j \mathscr{F}, \boldsymbol{P}(\cdot | \Delta_j))$. It follows from Theorem 3 that if $j \neq \infty$ and $\boldsymbol{P}(\Delta_j) > 0$ then the post-α_j process is a M.C. in $\boldsymbol{T}$ whose initial distribution is concentrated at the state j. On the other hand, if $\boldsymbol{P}(\Delta_\infty) > 0$ the post-α_∞ process has the initial distribution concentrated at the adjoined state ∞. For $t > 0$ let

$$r_{\infty j}(t) = \frac{r_j(\Delta; t) - \sum_{i \in \boldsymbol{I}} r_i(\Delta; 0)\, p_{ij}(t)}{\boldsymbol{P}(\Delta) - \sum_{i \in \boldsymbol{I}} r_i(\Delta; 0)}, \qquad j \neq \infty;$$

$r_{\infty\infty}(t) \equiv r_{i\infty}(t) \equiv 0$; $r_{ij}(t) \equiv p_{ij}(t)$, $i, j \in \boldsymbol{I}$. It is easily verified that (r_{ij}), $i, j \in \bar{\boldsymbol{I}}$, is a transition matrix which is measurable but not standard; and that the post-α_∞ process has a part of this matrix including the index ∞ as its transition matrix (cf. the remark at the end of § 4).

We observe that the properties of the M.C. expressed in Theorem 3 are extensions of the Markov property to which they reduce if α is a constant. Let us recall that our definitions of an optional α, the pre-α field and the post-α process and field are applicable without change to a general Borel measurable Markov process whose random variables take values (or, in the terminology used at the beginning of § 4, whose essential range is contained) in a Euclidean space or even an abstract metric space. Such a process $\{x_t, t \in \boldsymbol{T}\}$ will be said to have the *strong Markov property* iff for every optional α relative to $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$, the pre-$\alpha$ field and the post-α field are *conditionally independent relative to the initial random variable* ξ_0 of the post-α process, in the sense of (12), almost everywhere on the set where ξ_0 is finite; and that the post-α process is a Markov process whose transition probability functions are the same as the corresponding ones of the original process. According to this definition, a part of the content of Theorem 3 may be stated as follows:

If the M.C. $\{x_t, t\in \boldsymbol{T}\}$ has a discrete linear set as minimal state space compactified by the point at infinity ∞, and a standard transition matrix, then the version x_+ given in Theorem 7.1 has the strong Markov property.

Another part may be stated as follows:

If α is optional with domain of finiteness Δ, then for each $t>0$, the initial random variable $\xi_{\alpha+t}$ of the post-$(\alpha+t)$ process is finite a.e. on Δ. Thus the pre-$(\alpha+t)$ field and the post-$(\alpha+t)$ field are conditionally independent relative to $\xi_{\alpha+t}$, a.e. on Δ.

It should be remarked that the identification of the initial distribution of the post-α process with $\{r_j(\Lambda;0), j\in\boldsymbol{I}\}$, or more precisely the identification of $\boldsymbol{P}\{\xi(t,\omega)=j\,|\,\Lambda;\alpha\}$ with $r_j(\alpha,\alpha+t\,|\,\Lambda)$ given in Corollary 1 to Theorem 3, is not covered by the strong Markov property. This additional information will be needed in some applications. As an example let β be defined on Δ' and measurable $\mathscr{F}_\alpha$ such that $\alpha\leqq\beta<\infty$; then it is easy to see that β is optional in the same sense as α is. For $n\geqq 1$ define β_n on Δ' as follows:

$$\beta_n(\omega)=\alpha(\omega)+\frac{m+1}{n}\quad\text{if}\quad \alpha(\omega)+\frac{m}{n}\leqq\beta(\omega)<\alpha(\omega)+\frac{m+1}{n};$$

then $\beta_n(\omega)\downarrow\beta(\omega)$ and each β_n is likewise optional. By the strong Markov property, we have

$$\begin{aligned}\boldsymbol{P}&\left\{\beta_n(\omega)=\alpha(\omega)+\frac{m+1}{n};\ \xi\left(\frac{m+1}{n},\omega\right)=j\,\middle|\,\xi(0,\omega)=i\right\}\\&=\boldsymbol{P}\left\{\beta_n(\omega)=\alpha(\omega)+\frac{m+1}{n}\,\middle|\,\xi(0,\omega)=i\right\}p_{ij}\left(\frac{m+1}{n}\right).\end{aligned}$$

It follows that

$$\begin{aligned}&\boldsymbol{P}\{x(\beta_n(\omega),\omega)=j\,|\,x(\alpha(\omega),\omega)=i\}\\&\qquad=\boldsymbol{E}\{p_{ij}(\beta_n(\cdot)-\alpha(\cdot))\,|\,x(\alpha(\omega),\omega)=i\}.\end{aligned}$$

Letting $n\to\infty$, using a calculation similar to (8) on the left side and bounded convergence on the right, we conclude that the equation above remains true if β_n is replaced by β.

We proceed to discuss a special case of the strong Markov property where this property may be said to be at its strongest. The conditional independence relative to ξ_0 is to be replaced by absolute independence.

Theorem 4. A necessary and sufficient condition that the pre-α field $\mathscr{F}_\alpha$ and the post-α field $\mathscr{F}'_\alpha$ be independent with respect to $\boldsymbol{P}(\cdot\,|\,\Delta)$ is that for each $j\in\boldsymbol{I}$ there exists a function $\varrho_j(\cdot)$ on $\boldsymbol{T}^0$ such that for any $\Lambda\in\mathscr{F}_\alpha$ we have

$$r_j(s,t\,|\,\Lambda)=\varrho_j(t-s)\tag{14}$$

for a.a. $s \geqq 0$ in $A(\Lambda; \cdot)$ measure and $t > s$. If this is the case then we have in fact

$$(15) \qquad \varrho_j(\cdot) = \frac{r_j(\Delta; \cdot)}{\boldsymbol{P}(\Delta)}$$

so that $\varrho_j(\cdot)$ is continuous in $\boldsymbol{T}^0$ and can be extended to be continuous in $\boldsymbol{T}$.

Proof. Suppose that (14) holds, then it follows from (2) that

$$r_j(\Lambda; t) = \boldsymbol{P}(\Lambda)\, \varrho_j(t), \qquad t > 0.$$

Substituting into (5) and using the notation (7), we have

$$\boldsymbol{P}\{\Lambda; \xi(t_\nu, \omega) = j_\nu, 0 \leqq \nu \leqq N \mid \Delta\} = \boldsymbol{P}\{\Lambda \mid \Delta\}\, \varrho_{j_0}(t_0)\, Q.$$

Taking the particular case $\Lambda = \Delta$ and comparing the two equations we obtain

$$\begin{aligned} &\boldsymbol{P}\{\Lambda; \xi(t_\nu, \omega) = j_\nu, 0 \leqq \nu \leqq N \mid \Delta\} \\ &\qquad = \boldsymbol{P}\{\Lambda \mid \Delta\}\, \boldsymbol{P}\{\xi(t_\nu, \omega) = j_\nu, 0 \leqq \nu \leqq N \mid \Delta\}. \end{aligned}$$

This implies the asserted independence; the identification (15) and the continuity of $r_j(\Delta; \cdot)$ in $\boldsymbol{T}$ have already been noted above.

Conversely, suppose that there is the asserted independence. Let $u \geqq 0$ and put $\Lambda_u = \Lambda \cap \{\omega : \alpha(\omega) \leqq u\}$. It is clear that

$$A(\Lambda_u; s) = A\big(\Lambda; \min(s, u)\big).$$

Furthermore, by Corollary 1 to Theorem 3, we have

$$r_j(s, s+t \mid \Lambda) = \boldsymbol{P}\{\xi(t, \omega) = j \mid \Lambda; \alpha = s\}, \qquad t > 0, \text{ a.a. } s \geqq 0.$$

Using this interpretation we see that

$$r_j(s, s+t \mid \Lambda) = r_j(s, s+t \mid \Lambda_u), \qquad t > 0, \text{ a.a. } s \in [0, u].$$

It follows from (5) and these remarks that

$$\boldsymbol{P}\{\Lambda_u; \xi(t, \omega) = j\} = \int_{[0, u]} r_j(s, s+t \mid \Lambda)\, A(\Lambda; ds).$$

On the other hand, the assumed independence implies that

$$\begin{aligned} \boldsymbol{P}\{\Lambda_u; \xi(t, \omega) = j\} &= \boldsymbol{P}(\Lambda_u)\, \boldsymbol{P}\{\xi(t, \omega) = j \mid \Delta\} \\ &= \frac{1}{\boldsymbol{P}(\Delta)} \int_{[0, u]} r_j(\Delta; t)\, A(\Lambda; ds). \end{aligned}$$

Comparing the last two formulas we conclude, since u is arbitrary,

$$r_j(s, s+t \mid \Lambda) = \frac{1}{\boldsymbol{P}(\Delta)}\, r_j(\Delta; t), \qquad t > 0, \text{ a.a. } s \geqq 0.$$

By FUBINI's theorem, the last equation holds also if $0 \leqq s \notin Z$ and $0 \leqq t \notin Z_s$ where Z is of $A(\Lambda; \cdot)$ measure zero and each Z_s is of Lebesgue measure zero. Writing $s+t$ as t we obtain

$$r_j(s, t \mid \Lambda) = \frac{r_j(\Delta; t-s)}{\boldsymbol{P}(\Delta)} \tag{16}$$

if $0 \leqq s \notin Z$ and $0 < t-s \notin Z_s$. But for each $s \notin Z'$ where Z' is of $A(\Lambda; \cdot)$ measure zero, both members of (16) are continuous in $t \in (s, \infty)$ by Theorem 8.4; hence (16) is true if $0 \leqq s \notin Z \cup Z'$ and $t > s$. ▯

If in the statement and proof of Theorem 4 we consider only the set $\Lambda = \Delta$ we obtain the following result.

Corollary. The random variable α is independent of $\{\xi_t, t \in \boldsymbol{T}\}$ if and only if for each $j \in \boldsymbol{I}$ there exists a function $\varrho_j(\cdot)$ on $\boldsymbol{T}^0$ such that (14) holds for $\Lambda = \Delta$. In this case ϱ_j is given by (15).

The following particular case of Theorem 4 is so important in applications that we will state it as a separate theorem.

Theorem 5. Suppose that α is such that for a certain $j \neq \infty$, we have

$$\boldsymbol{P}\{\xi_0(\omega) = j \mid \Delta\} = 1. \tag{17}$$

Then the pre-α and post-α fields are independent. Furthermore for every $k \in \boldsymbol{I}$ and $\Lambda \in \mathscr{F}_\alpha$ we have

$$\boldsymbol{P}\{x(t, \omega) = k \mid \Lambda; \alpha = s\} = r_k(s, t \mid \Lambda) = p_{jk}(t-s) \tag{18}$$

for a.a. $s \geqq 0$ in $A(\Lambda; \cdot)$ measure and $t > s$.

Proof. Under the hypothesis (17) conditional independence relative to ξ_0, as given by the strong Markov property, becomes absolute independence, proving the first assertion. Hence Theorem 4 is applicable and we have from (14) and (15),

$$\begin{aligned} r_k(s, t \mid \Lambda) &= \boldsymbol{P}(\Delta)^{-1} r_k(\Delta; t-s) = \boldsymbol{P}(\Delta)^{-1} \boldsymbol{P}\{\Delta; \xi_0(\omega) = j; \xi_{t-s}(\omega) = k\} \\ &= \boldsymbol{P}(\Delta)^{-1} r_j(\Delta; 0) p_{jk}(t-s) = p_{jk}(t-s). \quad ▯ \end{aligned}$$

In this case the post-α process or the "future" is therefore absolutely independent of the "past", and "starts from scratch" at the state j.

Note that Theorem 8.2, our first result concerning optionality, is now subsumed under Theorem 5, the last one of a sequence of deductions, thus completing a circuit.

Notes. The main results of this section are contained in CHUNG [6], but the proofs of Theorems 2 and 3 have been considerably simplified and certain other points have been strengthened. A somewhat different formulation of the strong Markov property is given in YUŠKEVIČ [1] whose result in the case considered here is a little less precise than our Theorem 3 but includes the example involving β given before Theorem 4. For a discussion of strong Markov property for Markov processes with general state space see the last-cited paper and BLUMENTHAL [1], DYNKIN and YUŠKEVIČ [1]. Further literature and historical remarks can be found in the cited papers. For a second proof of part of Theorem 3 see AUSTIN [3]. The proof of Corollary 3 to Theorem 3 was communicated orally by AUSTIN; it can also be proved in a similar way as (11.17) below.

It will appear in subsequent applications (§§ 11, 14, 15) that the strong Markov property as defined here and similarly by the other authors is not by itself sufficient even for the simplest situation (e.g. in Theorem 15.2). The separability of the post-α process is essential and the continuity of the conditional probability $r_j(s, \cdot|\Lambda)$ will be needed (e.g. in Theorem 11.7). An ideal formulation should include these features. Finally, a more satisfactory approach to the theory should be obtained by an "intrinsic compactification" of the minimal state space by adjoining not a single undiscriminating ∞ but a set of ficticious states so as to restore (12) on the exceptional set $\{\omega: \xi_0(\omega) \notin \boldsymbol{I}\}$. This may also be formulated (cf. YUŠKEVIČ [1]) as the problem of "imbedding" a given M.C. in a Markov process with a larger, possibly nondenumerable, state space for which the strong Markov property embodied in (12) will hold without restriction on the value of ξ_0. In this direction see RAY [1] where unfortunately the application to the denumerable case is faulty and must be corrected. For the closely related theory of "boundaries" see FELLER [5] which is couched in more analytical terms.

Since the appearance of the first edition of the book, the strong Markov property has been extended to any exit time to an atomic boundary point. This is the starting point of the corresponding "boundary theory" developed in CHUNG [12] and [14].

§ 10. Classification of states

In this section the continuous parameter analogues of the main developments of §§ I.3—7 will be given. The corresponding discrete parameter results in Part I will be used; on the other hand the content of this section is independent of §§ II.5—9 except for some casual references to the beginning of § 5.

Associated with the c.p. M.C. $\mathfrak{C}=\{x_t, t\geqq 0\}$ is a one-parameter family of d.p. M.C.'s defined as follows. For each $h>0$ the d.p. stochastic process

$$\mathfrak{C}_h=\{x_{nh}, n\geqq 0\}$$

is clearly a d.p. M.C. with the same initial distribution as $\mathfrak{C}$ and the one-step transition matrix $(p_{ij}(h))$, $i, j \in \boldsymbol{I}$. Its n-step transition probabilities are given by $p_{ij}(nh)$. This d.p. M.C. will be called the *discrete skeleton of* $\mathfrak{C}$ *at the scale h*. The relations of $\mathfrak{C}_h$, $h > 0$, among themselves and to $\mathfrak{C}$ are often useful, as in the proof of Theorem 2.5, and will figure prominently in this section.

We begin by extending certain notions from the discrete parameter case. We write $i \rightsquigarrow j$ iff there exists a $t > 0$ such that $p_{ij}(t) > 0$; we say that i *communicates with* j and write $i \leftrightsquigarrow j$ iff $i \rightsquigarrow j$ and $j \rightsquigarrow i$. If the transition matrix is measurable, it follows from Theorem 1.4 that $i \leftrightsquigarrow i$ if $i \notin F$, where F is the class of states defined in Theorem 1.3. (We need not rule out the ficticious states here as we did in § 4, but they will play only a trivial role in the discussion.) It follows from the second assertion of Theorem 1.4 that if $i \rightsquigarrow j$ in $\mathfrak{C}$ then $i \rightsquigarrow j$ in each $\mathfrak{C}_h$; the converse is trivial. Thus the notion of communication in $\mathfrak{C}$ is equivalent to that in any $\mathfrak{C}_h$; in particular the classification of all states into classes of mutually communicating ones is identical in $\mathfrak{C}$ and all the $\mathfrak{C}_h$'s. Similarly the notion of *essentialness* is the same. However, if $i \leftrightsquigarrow i$ the period of i considered in $\mathfrak{C}_h$ is always equal to one, since $p_{ii}(nh) > 0$ for all n by Theorem 1.4. This is an important simplification for discrete skeletons regarded as d.p. M.C.'s. We note also the obvious significance of Theorem 1.5 (viz. Theorem 5.2) which is however not needed here.

The continuous parameter analogue of the main limit theorem of § I.6 follows easily from the discrete parameter theorem.

Theorem 1. Let (p_{ij}) be a measurable transition matrix. Then for every i and j, the following limit exists:

$$\lim_{t\to\infty} p_{ij}(t) = \pi_{ij}. \tag{1}$$

Proof. Suppose first that $i \notin F$; then its period in any $\mathfrak{C}_h$ is equal to one as just remarked. We have by Theorem I.6.4

$$\lim_{t\to\infty} p_{ij}(nh) = \pi_{ij}(h). \tag{2}$$

Hence given $\varepsilon > 0$, there exists n_0 such that $|p_{ij}(nh) - p_{ij}(n'h)| < \varepsilon/3$ if $n, n' > n_0$. Since p_{ij} is uniformly continuous in $\boldsymbol{T}$ by the Corollary to Theorem 2.2 we may choose h so that the oscillation of p_{ij} in any interval of length h is less than $\varepsilon/3$. It follows that $|p_{ij}(t) - p_{ij}(t')| < \varepsilon$ if $t, t' > n_0 h$ proving the existence of the limit in (1).

Now let $i \in F$ and recall Theorem 1.3. The transition matrix (Π_{IJ}), is standard, hence $\lim_{t\to\infty} \Pi_{KJ}(t)$ exists by what has just been proved. It follows then from the relations in (1.16) and (1.17) that

$$\lim_{t\to\infty} p_{ij}(t) = \lim_{t\to\infty} \Pi_{iJ}(t)\, u_j = \sum_K \Pi_{iK}(s) \lim_{t\to\infty} \Pi_{KJ}(t)\, u_j$$

exists, where $s > 0$ is arbitrary. ▯

Among the analogues of the discrete parameter theorems in § I.7 we prove only the following one.

Theorem 2. We have for any $s>0$,

$$\pi_{ij}=\sum_k \pi_{ik}\, p_{kj}(s)=\sum_k p_{ik}(s)\,\pi_{kj}=\sum_k \pi_{ik}\,\pi_{kj}. \tag{3}$$

The matrix (π_{ij}) satisfies the conditions (1.1), (1.2) and (1.3′). If $\pi_{ii}\neq 0$, then $\sum_j \pi_{ij}=1$.

Proof. We have by (1.C), Theorem 1 and Fatou's lemma,

$$\pi_{ij}\geqq \sum_k \pi_{ik}\, p_{kj}(s). \tag{4}$$

Summing over j we have

$$\sum_j \pi_{ij}\geqq \sum_k \pi_{ik}\sum_j p_{kj}(s)=\sum_k \pi_{ik}.$$

It follows that equality must hold in (4). This proves the first equation in (3); the second follows from the uniform convergence of the series in (1.C) with respect to t. By Fatou's lemma again we have $\sum_k \pi_{ik}\leqq 1$. Hence the series on the right side of (4) converges uniformly with respect to s in $(0,\infty)$. Letting $s\to\infty$ in (4) with equality we obtain the third equation in (3). We have thus proved that the matrix (π_{ij}) satisfies the stated conditions for (u_{ij}) in Theorem 1.2. The last assertion then follows from (1.9). ▯

We now come to the notion of recurrence. In the rest of this section the transition matrix (p_{ij}) will be assumed to be standard and the M.C. $\{x_t, t\in \boldsymbol{T}\}$ well-separable and measurable with the state space as specified at the beginning of § 5.

Theorem 3. For any $i\neq\infty$ we have

$$\{\omega\colon S_i(\omega) \text{ is unbounded}\}\doteq\{\omega\colon \mu[S_i(\omega)]=\infty\}. \tag{5}$$

Proof. Let Δ be the set on the left side of (5). For a.a. $\omega\in\Delta$, we have $x(r,\omega)=i$ on an unbounded set of $r\in R$ where R is a separability set. It follows from Theorem 5.3 or the remarks preceding it that given any $\eta>0$ there exists a positive $\varepsilon(\eta)$ independent of r such that

$$\boldsymbol{P}\left\{\mu[S_i(\omega)\cap(r,r+\varepsilon)]>\frac{\varepsilon}{2}\,\middle|\, x(r,\omega)=i\right\}\geqq 1-\eta.$$

Hence we have for every $r\in R$,

$$\begin{aligned}
&\boldsymbol{P}\left\{\mu[S_i(\omega)\cap(r,\infty)]>\frac{\varepsilon}{2}\right\}\\
&\quad\geqq \lim_{n\to\infty}\sum_{m=0}^{\infty}\boldsymbol{P}\left\{x\left(r+\frac{\nu}{2^n},\omega\right)\neq i,\ 0<\nu<m;\ x\left(r+\frac{m}{2^n},\omega\right)=i\right\}\times\\
&\quad\times\boldsymbol{P}\left\{\mu\left[S_i(\omega)\cap\left(r+\frac{m}{2^n},\ \infty\right)\right]>\frac{\varepsilon}{2}\,\middle|\, x\left(r+\frac{m}{2^n},\omega\right)=i\right\}\geqq \boldsymbol{P}(\Delta)(1-\eta).
\end{aligned}$$

Letting $r \to \infty$ we see that

$$\boldsymbol{P}\{\mu[S_i(\omega)] = \infty\} \geqq \boldsymbol{P}(\Delta)(1-\eta).$$

Since η is arbitrary we conclude that the probability of the left member of (5) does not exceed that of the right member. Since the former set contains the latter we have proved (5). ▯

Theorem 4. For any $i \neq \infty$, we have

$$(6)\quad \boldsymbol{P}\{S_i(\omega) \text{ is unbounded} \mid x(0,\omega)=i\} = \boldsymbol{P}\{\mu[S_i(\omega)]=\infty \mid x(0,\omega)=i\}$$

and either probability is equal to zero or one according as

$$(7)\qquad \boldsymbol{E}\{\mu[S_i(\omega)] \mid x(0,\omega)=i\} = \int_0^\infty p_{ii}(t)\,dt$$

is finite or infinite. Furthermore, the integral in (7) diverges if and only if

$$(8)\qquad \sum_{n=0}^{\infty} p_{ii}(nh)$$

diverges for some $h > 0$, in which case it diverges for every $h > 0$.

Proof. The equation (6) follows from (5) and the equation (7) follows from Fubini's theorem as in Theorem 5.1. Next, let

$$\min_{0 \leqq r \leqq h} p_{ii}(r) = \delta(h) > 0.$$

It follows from the properties of the transition matrix that $\lim_{n \to 0} \delta(h) = 1$ and

$$\min_{0 \leqq r \leqq h} p_{ii}(t+r) \geqq p_{ii}(t) \min_{0 \leqq r \leqq h} p_{ii}(r) = p_{ii}(t)\,\delta(h);$$

consequently

$$m_n(h) = \min_{nh \leqq t \leqq (n+1)h} p_{ii}(t) \geqq \delta(h)\, p_{ii}(nh).$$

Similarly we have

$$M_n(h) = \max_{nh \leqq t \leqq (n+1)h} p_{ii}(t) \leqq \delta(h)^{-1} p_{ii}((n+1)h).$$

Combining these two inequalities we obtain,

$$(9)\qquad \left\{ \begin{aligned} \delta(h)\, h \sum_{n=0}^{N-1} p_{ii}(nh) &\leqq h \sum_{n=0}^{N-1} m_n(h) \leqq \int_0^{Nh} p_{ii}(t)\,dt \\ &\leqq h \sum_{n=0}^{N-1} M_n(h) \leqq \delta(h)^{-1} h \sum_{n=1}^{N} p_{ii}(nh). \end{aligned} \right.$$

Letting $N \to \infty$ we see that the integral in (7) and the series in (8) diverge together. This being true for each h the last assertion of Theorem 4 is proved.

Now suppose that the left member of (7) is finite; then clearly the right member of (6) is equal to zero. On the other hand, if the right member of (7) diverges, then so does the series in (8) for every $h>0$. By Theorem I.5.4 the state i is then recurrent in $\mathfrak{C}_1$ so that

$$\boldsymbol{P}\{x(n,\omega)=i \text{ for infinitely many values of } n \mid x(0,\omega)=i\}=1 .$$

A fortiori the left member of (6) is equal to one. Theorem 4 is completely proved. □

As a consequence of Theorem 4 we define recurrence as follows. The state i is *recurrent* or *nonrecurrent* according as either probability in (6) is equal to one or zero. Thus, i is recurrent if and only if $\int_0^\infty p_{ii}(t)\,dt =\infty$; or if and only if it is recurrent in some (and hence every) $\mathfrak{C}_h$. These two criteria could be used as alternative definitions of recurrence; they have the advantage of being immediately generalizable to a measurable (instead of standard) transition matrix. For, upon reexamination of the preceding proof we see that the condition $\varliminf_{h\to 0}\delta(h)>0$ is sufficient for the co-divergence of the integral in (7) and the series in (8). This condition is satisfied if $i\notin F$, by Theorem 1.3, while if $i\in F$ both series and integral are equal to zero. However, without a standard transition matrix the probability meaning of recurrence as given in Theorem 4 is lost.

The state i is called *absorbing* iff $p_{ii}(t)\equiv 1$ for all $t>0$. A necessary and sufficient condition for this is $q_i=0$, by (3.2). The following corollaries to Theorem 4 are immediate.

Corollary 1. If i is stable but not absorbing, then i is recurrent or nonrecurrent according as the number of i-intervals of $x(\cdot,\omega)$ is infinite or finite for a.a. ω in the set $\{\omega\colon x(0,\omega)=i\}$.

Corollary 2. For any $j\neq\infty$ we have

$$(10)\quad \boldsymbol{P}\{S_i(\omega) \text{ is unbounded } \mid x(0,\omega)=j\}=\boldsymbol{P}\{\mu[S_i(\omega)]=\infty \mid x(0,\omega)=j\};$$

and either probability is zero or one according as $\int_0^\infty p_{ji}(t)\,dt$ is finite or infinite. If i is recurrent and $i\rightsquigarrow j$ then $i\leftrightsquigarrow j$ and both probabilities in (10) are equal to one; if i is nonrecurrent then for every j both probabilities are equal to zero.

It follows furthermore that the classification of states and classes into recurrent and nonrecurrent ones is the same in $\mathfrak{C}$ as in each $\mathfrak{C}_h$. In a recurrent class each $S_i(\omega)$, $i\neq\infty$, is unbounded and has infinite measure; in a nonrecurrent class each $S_i(\omega)$ is bounded for a.a. ω. An inessential state [or class] is nonrecurrent.

A recurrent state i is called *positive* or *null* according as $\pi_{ii} > 0$ or $\pi_{ii} = 0$ in (1). The notion of *positiveness* in $\mathfrak{C}$ is equivalent to that in any $\mathfrak{C}_h$ and so is the resulting classification.

Notes. In this and the next four sections we do not propose to extend all the corresponding results in Part I, but only those which are either basic or offer something new.

Theorem 1 for $i \notin F$ is proved by LÉVY [1]. In connection with Theorem 2 we omit a discussion of the limit matrix at infinity analogous to Theorem 1.4; see DOOB [1] and KENDALL and REUTER [2]. Theorems 3 and 4 are from CHUNG [7]. It is curious to note that if p_{ii} is merely assumed to be continuous in $[0, \infty)$ and bounded at zero, then the divergence of the integral in (7) implies the divergence of the series in (8) except for a set of h of the first category; see Amer. Math. Monthly, 64 (1957), 120, the second solution (the first being incorrect). A special case of Corollary 2 to Theorem 4 where the sample functions are assumed to be step functions is discussed in URBANIK [1]. For extension of the notion of repetitive pattern in § I.8 to continuous time. see KINGMAN [1].

§ 11. Taboo probability functions

We proceed to develop the continuous parameter analogue of the theory of taboo probabilities given in § I.9. While the intuitive content is easily seen the formal details must be treated with caution and the main result (Theorem 8 below) depends essentially on the strong Markov property discussed in § 9 and its consequences given in Theorems 6 and 7 here. Some interesting properties of an instantaneous state are obtained as by-product which will perhaps play a role in a further study.

The M.C. $\{x_t, t \in \boldsymbol{T}\}$ will be assumed to be Borel measurable and well-separable with a standard transition matrix. It is not always necessary to specify more, but at a later point the version x_+ will be used to invoke the strong Markov property. An unsepcified state denoted by i, j, or k below belongs to $\boldsymbol{I}$, thus $\neq \infty$.

On the set $\Delta_i = \{\omega: x_0(\omega) = i\}$ with $\boldsymbol{P}(\Delta_i) > 0$ we define

$$(1) \qquad \varrho_i(\omega) = \inf\{t: t > 0;\ x(t, \omega) \neq i\}.$$

By Theorem 5.5, ϱ_i is a random variable which is finite a.e. on Δ_i, and

$$(2) \qquad \boldsymbol{P}\{\varrho_i(\omega) \geqq t \mid \Delta_i\} = e^{-q_i t}, \qquad t \geqq 0.$$

In particular if i is instantaneous then $\boldsymbol{P}\{\varrho_i(\omega) = 0 \mid \Delta_i\} = 1$. Under the hypothesis that i is the initial state, ϱ_i is the *first exit time from* i and also the *first sojourn time in* i.

Let H be an arbitrary, possibly empty set, (in which case H is simply omitted from the notation) of states excluding the adjoined state ∞, to be called the *taboo set*. We define for $t>0$,

$$(3)\qquad {}_Hp_{ij}(t)=\boldsymbol{P}\{x(s,\omega)\notin H,\ \min[t,\varrho_i(\omega)]<s<t;\ x(t,\omega)=j\,|\,\Delta_i\}$$

where, as in similar instances below, the notation means: $x(s,\omega)\notin H$ for all s in $(\varrho_i(\omega),t)$ provided $\varrho_i(\omega)<t$, otherwise this condition is to be omitted. It will be shown shortly (in the proof of Theorem 1) that the ω-set whose probability is given in (3) is indeed measurable $\mathscr{F}$. Observe that if i is instantaneous or if i is stable and $i\notin H$ we can replace the min in (3) by 0. The case where i is stable and $i\in H$ frequently calls for a special consideration. In general we may write more explicitly:

$$(4)\qquad \begin{cases} {}_Hp_{ij}(t)=\delta_{ij}\,e^{-q_it}+ \\ \qquad\qquad +\boldsymbol{P}\{\varrho_i(\omega)<t;\ x(s,\omega)\notin H,\ \varrho_i(\omega)<s<t;\ x(t,\omega)=j\,|\,\Delta_i\}\end{cases}$$

where we agree that $e^{-\infty}=0$ as usual. The probability ${}_Hp_{ij}(t)$ is called the *transition probability from i to j in time t under the taboo H*; in case i is stable and $i\in H$ the taboo is imposed only after the first exit from i. We note that if $i\neq j$ then ${}_jp_{ij}(t)\equiv 0$ by Theorem 6.3, *a fortiori* ${}_Hp_{ij}(t)\equiv 0$ if $j\in H$. On the other hand if i is stable and $i\in H$ we have ${}_Hp_{ii}(t)\equiv e^{-q_it}$, since the second term on the right side of (4) then does not exceed $\boldsymbol{P}\{t\in S_i(\omega)-S_i^-(\omega)\}=0$ by Theorem 6.3.

We have, taking account of the preceding remarks,

$$(5)\qquad {}_Hp_{ij}(s+t)=\sum_k {}_Hp_{ik}(s)\,{}_Hp_{kj}(t),\qquad s>0,\ t>0;$$

where except in the case i is stable and $i\in H$ the summation may be taken over $k\notin H$ instead. Let further

$$(6)\qquad {}_HD_i(t)=\sum_j {}_Hp_{ij}(t)=\boldsymbol{P}\{x(s,\omega)\notin H,\ \min[t,\varrho_i(\omega)]<s<t\,|\,\Delta_i\}.$$

It is clear that $({}_Hp_{ij})$, $i,j\in\boldsymbol{I}$, is a substochastic transition matrix discussed at the end of § 3. From this and the results of § 1 we may draw several conclusions.

Theorem 1. Each ${}_Hp_{ij}$ is a continuous function in $(0,\infty)$.

Proof. According to the remarks following Theorem 3.3 it is sufficient to prove that each ${}_Hp_{ij}$ is measurable. Without loss of generality we may suppose that $\boldsymbol{P}(\Delta_i)=1$. Define $\xi_k(t,\omega)=1$ or 0 according as $x(t,\omega)=k$ or not; define also

$$e(t,r,\omega)=\begin{cases}1 & \text{if } \varrho_i(\omega)<r<t;\\ 0 & \text{otherwise.}\end{cases}$$

For each r, we have

$$\{(t,\omega): e(t,r,\omega)=1\}=\{t: t>r\}\times\{\omega: \varrho_i(\omega)<r\}\in\mathscr{B}\times\mathscr{F}.$$

Putting

$$\eta_k(t,\omega)=\sum_{r\in R} e(t,r,\omega)\,\xi_k(r,\omega)$$

where R is a separability set, we see that

$$\begin{aligned}\{(t,\omega): x(s,\omega)\notin H, \min[t,\varrho_i(\omega)]<s<t;\ x(t,\omega)=j\}\\ =\{(t,\omega): \sum_{k\in H}\eta_k(t,\omega)=0;\ x(t,\omega)=j\}.\end{aligned}$$

Obviously η_k is measurable $\mathscr{B}\times\mathscr{F}$; hence the set on the left belongs to $\mathscr{B}\times\mathscr{F}$. Hence ${}_Hp_{ij}(t)$, being the $\boldsymbol{P}$-measure of the section of this (t,ω) set for a fixed t, is a Borel measurable function of t. ▯

Corollary. ${}_HD_i$ is a continuous function in $(0,\infty)$. This follows from the Corollary to Theorem 1.3 which implies the uniform convergence with respect to t of the series in (6) in every finite interval.

We have ${}_Hp_{ij}(t)\leqq p_{ij}(t)$ and if i is stable then ${}_Hp_{ii}(t)\geqq e^{-q_i t}$ by (4) so that $\lim_{t\to 0}{}_Hp_{ij}(t)=\delta_{ij}$. The same conclusion holds if i is instantaneous and H is finite by Theorem 6.2 (ii). For an instantaneous i the general situation is as follows.

Theorem 2. For any H and an instantaneous i,

$$\lim_{t\to 0}{}_Hp_{ii}(t)=\lim_{t\to 0}{}_HD_i(t) \tag{7}$$

exists and is equal to zero or one. If it is zero then ${}_Hp_{ij}(t)\equiv{}_Hp_{ji}(t)\equiv 0$ for every j and $t>0$. If it is one then ${}_Hp_{ij}$ is uniformly continuous in $(0,\infty)$ for each j.

Proof. It is obvious from (6) that ${}_HD_i(t)$ is nondecreasing as $t\downarrow 0$; furthermore

$${}_Hp_{ii}(t)\leqq{}_HD_i(t)\leqq{}_Hp_{ii}(t)+\sum_{j\neq i}p_{ij}(t),$$

from which (7) follows. Thus the limits $u_{ij}=\lim_{t\to 0}{}_Hp_{ij}(t)$ exist and they satisfy the conditions (1.1) to (1.3) by (5); these also follow from the fact that $({}_Hp_{ij})$ is a substochastic transition matrix. Using the notation of Theorem 1.3, if $j\in J\neq F$ and $i\in J$ we have $u_{ij}=u_j>0$ by (1.5). But if $i\neq j$ we have $u_{ij}=0$ as already noted; thus we must have $i=j$ and each $J\neq F$ consists of one element only. Furthermore it follows from (1.6) that $u_{ii}=1$ if $i\notin F$. We have therefore proved that the common value of the limits in (7) is either zero or one. If it is zero then $i\in F$ and so ${}_Hp_{ji}(t)\equiv 0$ by (1.13); on the other hand we have ${}_Hp_{ij}(t)\leqq {}_HD_i(t)\equiv 0$. If the first limit in (7) is one then each ${}_Hp_{ij}$ is uniformly continuous as in Theorem 2.2. ▯

The meaning of the limits in (7) is given in the next theorem.

Theorem 3. Let i be instantaneous and $i\notin H$, then we have for each t,

$$\text{(8)}\qquad \begin{cases} \boldsymbol{P}\left\{\bigcup_{k\in H} S_k(\omega)\cap(t,t+\varepsilon)=0 \text{ for some } \varepsilon>0 \mid x(t,\omega)=i\right\} \\ =\boldsymbol{P}\left\{\bigcup_{k\in H} S_k(\omega)\cap(t-\varepsilon,t)=0 \text{ for some } \varepsilon>0 \mid x(t,\omega)=i\right\} \\ =\lim\limits_{\varepsilon\to 0} {}_Hp_{ii}(\varepsilon). \end{cases}$$

Proof. The first probability in (8) is equal to $\lim\limits_{\varepsilon\to 0} {}_HD_i(\varepsilon)$. The second probability is the limit as $\varepsilon\downarrow 0$ of

$$\text{(9)}\qquad p_i(t)^{-1}\sum_j p_j(t-\varepsilon)\,{}_Hp_{ji}(\varepsilon).$$

If $\lim\limits_{\varepsilon\to 0} {}_Hp_{ii}(\varepsilon)=0$ then (9) is equal to zero identically in ε by Theorem 2; if $\lim\limits_{\varepsilon\to 0} {}_Hp_{ii}(\varepsilon)=1$ then (9) is not less than $p_i(t)^{-1}p_i(t-\varepsilon)\,{}_Hp_{ii}(\varepsilon)$ and hence approaches one. ▯

Observe that for a stable i the first two probabilities in (8) are equal to zero or one according as $i\notin H$ or $i\in H$. For each $i\in\boldsymbol{I}$, a subset H of $\boldsymbol{I}$ will be called *adjacent* or *nonadjacent* to i according as both these probabilities are equal to zero or one. It follows from the preceding remark and Theorem 3 that this is a true dichotomy. The complement of a set nonadjacent to i will be called a *fine neighborhood* of i. It follows from Theorem 6.2 that each fine neighborhood contains i; furthermore the system of fine neighborhoods of all states defines a topology on $\boldsymbol{I}$ called the *fine topology*. The next theorem characterizes this topology by means of the transition probabilities and gives its basic property.

Theorem 4. A base of neighborhoods for i is given by the sets $\{B_i(\delta,\delta'),\ \delta>0,\ \delta'>0\}$, or alternatively by $\{C_i(\delta),\ \delta>0\}$, defined below:

$$B_i(\delta,\delta')=\{k\in\boldsymbol{I}\colon p_{ki}(\delta)>1-\delta'\};$$

$$C_i(\delta)=\{k\in\boldsymbol{I}\colon \sup_{0\leqq t\leqq\delta} p_{ki}(t)>1-\delta\}.$$

Two distinct states i and j have two disjoint fine neighborhoods; namely, $\boldsymbol{I}$ endowed with the fine topology is a Hausdorff space. Furthermore, almost every sample function $x_+(\cdot,\omega)$, $[x_-(\cdot,\omega)]$ is right [left] continuous in the fine topology on the set $\{t\colon x(t,\omega)\in\boldsymbol{I}\}$.

Proof. The equivalence of the two bases follows from the inequalities:

$$p_{ki}(\delta) \leqq \sup_{0 \leqq t \leqq \delta} p_{ki}(t) \leqq p_{ki}(\delta) \left[\inf_{0 \leqq t \leqq \delta} p_{ii}(t)\right]^{-1}.$$

The Hausdorff separation property is demonstrated by the obvious relation

$$B_i(\delta, \tfrac{1}{2}) \cap B_j(\delta, \tfrac{1}{2}) = 0,$$

since $p_{ki}(\delta) + p_{kj}(\delta) \leqq 1$ for each k and $\delta > 0$. It remains to prove that the sets $C_i(\delta)$ form a base of fine neighborhoods for i and to prove the last assertion of the theorem for x_+ (that for x_- being entirely similar). Both will be accomplished if we prove that given any i and $C_i(\delta_0)$, for a.e. ω the sample function $x(\cdot, \omega)$ has the following property: if $x(t, \omega) = i$, then there exists $h(\omega) > 0$ such that

$$(10) \qquad x(s, \omega) \in C_i(\delta_0) \cup \{\infty\} \quad \text{for} \quad s \in \big(t, t + h(\omega)\big).$$

This is indeed trivial if i is stable, by Theorem 5.7. In general we consider the martingale $\{\eta_t^A, 0 \leqq t < A\}$ introduced in the proof of Theorem 6.1. It follows from the discussion there that for a.e. ω, and every generic t, we have

$$\lim_{r \in R, r \downarrow t} p_{x(r,\omega), j}(A - r) = \eta^A(t, \omega)$$

where R is a separability set for $\{x_t\}$. If $x_+(t, \omega) = i$, then as $r \downarrow t$ there are infinitely many values of r for which $x(r, \omega) = i$, and consequently by the continuity of p_{ij} the limit on the left side above must be $p_{ij}(A - t)$. Thus we have in particular for $i = j$

$$(11) \qquad \lim_{r \in R, r \downarrow t} p_{x(r,\omega), i}(A - r) = p_{ii}(A - t).$$

Fix a δ such that $\delta < \delta_0$ and $\inf_{0 \leqq t \leqq \delta} p_{ii}(t) > 1 - \delta_0$. For t in $S_i(\omega) \cap [(m-1)\delta, m\delta]$ where m is a positive integer, apply (11) with $A = m\delta$. It follows that for a.e. ω we have

$$\varliminf_{r \in R, r \downarrow t} \left[\sup_{0 \leqq u \leqq \delta} p_{x(r,\omega), i}(u)\right] \geqq \inf_{0 \leqq t \leqq \delta} p_{ii}(t).$$

Hence for a.e. ω there exists $h(\omega) > 0$ such that

$$x(r, \omega) \in C_i(\delta_0) \quad \text{for} \quad r \in R \cap \big(t, t + h(\omega)\big)$$

for every t in the said range. This relation implies the desired relation (10) by separability and by varying m. ▯

Let us note that the unilateral continuity asserted in Theorem 4 amounts to a strengthening of Theorem 3 from a "fixed t" version to a "generic t" version. Needless to add that a bilateral strengthening would be out of question, e.g., at a t in $\overline{S_i(\omega)} \cap \overline{S_j(\omega)}$ where i is instantaneous and j is stable.

Corollary. If i is instantaneous and H is adjacent to i, then there exists an infinite sequence $\{i_\nu\}$ from H such that for every $j \in \boldsymbol{I}$ and $t > 0$ we have

$$\lim_{\nu\to\infty} p_{i_\nu j}(t) = p_{ij}(t).$$

Proof. Since H is adjacent to i it intersects every fine neighborhood of i and so in particular every $B_i(\delta, \delta')$. Taking two sequences $\delta_n \downarrow 0$ and $\delta'_n \downarrow 0$ it follows that there exists a sequence $\{i_\nu\}$ from H such that

$$p_{i_\nu i}(\delta_n) > 1 - \delta'_n.$$

Now we write for $\delta_n < t$:

$$p_{i_\nu j}(t) = p_{i_\nu i}(\delta_n)\, p_{ij}(t - \delta_n) + \sum_{k \neq i} p_{i_\nu k}(\delta_n)\, p_{kj}(t - \delta_n),$$

and the corollary follows at once from the continuity of p_{ij}. ▯

Turning to the behavior of ${}_H p_{ij}(t)$ as $t \to \infty$, we prove the following analogue of Theorem I.9.3. We write as in the discrete parameter case $j \rightsquigarrow H$ iff there is a state $k \in H$ such that $j \rightsquigarrow k$.

Theorem 5. If $j \rightsquigarrow H$ then

$$\int_0^\infty {}_H p_{ij}(t)\, dt < \infty.$$

Proof. Let $k \in H$ and $p_{jk}(u) > 0$, $u > 0$. For $t > 0$ and $n \geqq 0$ we have from (4),

$$\begin{aligned} {}_H p_{ij}(nu+t)\, p_{jk}(u) \leqq e^{-q_i(nu+t)} + \boldsymbol{P}\{\varrho_i(\omega) \leqq nu+t;\\ x(s,\omega) \notin H,\ \varrho_i(\omega) < s < nu+t;\ x((n+1)u+t, \omega) = k \mid \Delta_i\} \end{aligned}$$

For fixed u let $\Lambda_n(t)$ be the ω-set whose conditional probability is given by the last term above; then if $\nu < n$ we have

$$\begin{aligned} \Lambda_\nu(t) \cap \Lambda_n(z) \subset \{\omega\colon \varrho_i(\omega) \leqq \nu u + t;\\ x(s,\omega) \notin H,\ \varrho_i(\omega) < s < nu+t;\ x((\nu+1)u+t, \omega) = k\}. \end{aligned}$$

Since $k \in H$ the set on the right side is empty if $\nu \leqq n-2$, and is a null set if $\nu = n-1$ by Theorem 6.3. Hence the sets $\{\Lambda_n(t), n \geqq 0\}$ are disjoint up to a null set. It follows from the above that

$$p_{jk}(u) \int_0^u {}_H p_{ij}(nu+t)\, dt \leqq \int_0^u e^{-q_i(nu+t)}\, dt + \int_0^u \boldsymbol{P}\{\Lambda_n(t) \mid \Delta_i\}\, dt.$$

Summing over n we obtain

$$p_{jk}(u) \int_0^\infty {}_H p_{ij}(t)\, dt \leqq \int_0^\infty e^{-q_i t}\, dt + \int_0^u 1\, dt = \frac{1}{q_i} + u. \ ▯$$

We now define on Δ_i new random variables as follows:

$$\alpha_{ij}(\omega) = \inf\{t\colon t > \varrho_i(\omega);\ x(t,\omega) = j\}. \tag{12}$$

Note that for a.a. ω, $\alpha_{ii}(\omega) > \varrho_i(\omega)$ if i is stable, and $\alpha_{ii}(\omega) = \varrho_i(\omega) = 0$ if i is instantaneous. Furthermore we define

$$_H\alpha_{ij}(\omega) = \begin{cases} \alpha_{ij}(\omega), & \text{if} \quad \alpha_{ij}(\omega) \leqq \inf\limits_{k\in H} \alpha_{ik}(\omega); \\ \infty, & \text{otherwise.} \end{cases}$$

Thus $_H\alpha_{ij} \equiv \alpha_{ij}$ if H is empty. The domain of finiteness of $_H\alpha_{ij}$ will be denoted by $_H\Delta_{ij}$. It is easy to verify that $_H\alpha_{ij}$ is a random variable; and for each $t > 0$, $\boldsymbol{P}\{_H\alpha_{ij}(\omega) = t\} = 0$ by Theorem 6.3. Furthermore we have $\{\omega\colon {_H\alpha_{ij}(\omega)} < t\} \in \mathscr{F}\{x_s,\ s \leqq t\}$ so that $_H\alpha_{ij}$ is optional relative to $\{x_t,\ t \in \boldsymbol{T}\}$. Let $_HF_{ij}$ be its distribution function, namely

$$_HF_{ij}(t) = \boldsymbol{P}\left\{\alpha_{ij}(\omega) \leqq \min\left[\inf_{k\in H} \alpha_{ik}(\omega),\ t\right] \Big|\, \Delta_i\right\}$$

or equivalently

$$_HF_{ij}(t) = \boldsymbol{P}\{\alpha_{ij}(\omega) \leqq t;\ x(s,\omega) \notin H,\ \min[t, \varrho_i(\omega)] < s < \alpha_{ij}(\omega) \,|\, \Delta_i\}. \tag{13}$$

In particular

$$F_{ij}(t) = \boldsymbol{P}\{\alpha_{ij}(\omega) \leqq t \,|\, \Delta_i\}.$$

If $i \neq j$, $_H\alpha_{ij}$ is called the *first entrance time from i to j under the taboo H*; if i is stable $_H\alpha_{ii}$ is the *first return time to i* (after leaving i at the first exit time ϱ_i) *under the taboo H*. Note that $F_{ij}(0) = 0$ or δ_{ij} according as i is stable or instantaneous. Furthermore if $j \notin H$, and either j is stable or H is finite then $\boldsymbol{P}\{\alpha_{ij}(\omega) = \inf_{k\in H} \alpha_{ik}(\omega) \,|\, \Delta_i\} = 0$ by Theorem 7.4. On the other hand, if j is instantaneous it can be shown that the last written probability is equal to one or zero according as H is adjacent or nonadjacent. Next, we have for each $t > 0$,

$$\boldsymbol{P}\left\{\inf_{k\in H} \alpha_{ik}(\omega) = t \,\Big|\, \Delta_i\right\} = 0. \tag{14}$$

To prove (14) we show that for each j,

$$\boldsymbol{P}\left\{\inf_{k\in H} \alpha_{ik}(\omega) = t;\ x(t,\omega) = j \,\Big|\, \Delta_i\right\} = 0.$$

This is trivial if j is stable or if j is instantaneous and H is nonadjacent to j. If j is instantaneous and H is adjacent to j it follows from the second equation in (8). Finally we remark that

$$_Hp_{ij}(t) = \boldsymbol{P}\left\{t < \inf_{k\in H} \alpha_{ik}(\omega);\ x(t,\omega) = j \,\Big|\, \Delta_i\right\} \tag{15}$$

where the "$<$" may be replaced by "$\leqq$" on account of (14).

Theorem 6. Two separable and Borel measurable M.C.'s $\{x_t\}$ and $\{\hat{x}_t\}$ having the same transition matrix have the same ${}_Hp_{ij}$ and ${}_HF_{ij}$.

Proof. We show this by expressing ${}_Hp_{ij}(t)$ in terms of the transition matrix. Let us evaluate the second term on the right side of (4) when i is stable, the other case being trivial. The argument is typical in analysing separable sample functions. Let R be enumerated and let $\{s_\nu^{(n)}, 1 \leqq \nu \leqq n\}$ be the first n elements of R that belong to $[0, t)$, in increasing order. Let

$$\Gamma = \Delta_i \cap \{\omega: 0 < \varrho_i(\omega) < t;\ x(s, \omega) \notin H,\ \varrho_i(\omega) < s < t;\ x(t, \omega) = j\},$$

$$\Gamma_n = \Delta_i \cap \bigcup_{m=1}^{n} \{\omega: x(s_\nu^{(n)}, \omega) = i,\ 1 \leqq \nu < m;\ x(s_m^{(n)}, \omega) \neq i;$$
$$x(s_\nu^{(n)}, \omega) \notin H,\ m \leqq \nu \leqq n;\ x(t, \omega) = j\}.$$

Put $s_0^{(n)} = 0$ and $\delta^{(n)} = \max\limits_{1 \leqq \nu \leqq n} (s_\nu^{(n)} - s_{\nu-1}^{(n)})$. If $\omega_0 \in \Gamma$, then as soon as $\delta^{(n)} < \min[\varrho_i(\omega_0), t - \varrho_i(\omega_0)]$ we have clearly $\omega_0 \in \Gamma_n$; hence $\Gamma \subset \liminf\limits_n \Gamma_n$. On the other hand if $\omega_1 \in \limsup\limits_n \Gamma_n$ and $x(\cdot, \omega_1)$ is separable (with respect to R) in $\boldsymbol{T}$, then clearly $\varrho_i(\omega_1) < t$ and $x(t, \omega_1) = j$. Suppose that $x(s, \omega_1) \in H$ for some $s \in (\varrho_i(\omega_1), t)$, then the same is true for a certain $s_{\nu_0}^{(n_0)} > \varrho_i(\omega_1)$. For each n let $m = m(n)$ be the least index such that $x(s_m^{(n)}, \omega_1) \neq i$. Since $s_m^{(n)} \downarrow \varrho_i(\omega_1)$ as $n \to \infty$ by separability, there exists an n_1 such that if $n \geqq n_1$ then $s_m^{(n)} < s_{\nu_0}^{(n_0)}$. Thus $\omega_1 \notin \bigcup\limits_{n=n_1}^{\infty} \Gamma_n$, contradicting the choice of ω_1. Hence $\omega_1 \in \Gamma \cup [\Delta_i \cap \{\omega: \varrho_i(\omega) = 0\}]$ and so $\limsup\limits_n \Gamma_n \subset \Gamma$ since $\boldsymbol{P}\{\Delta_i; \varrho_i(\omega) = 0\} = 0$ when i is stable. We have therefore proved that

$$\boldsymbol{P}(\Gamma) = \lim_{n \to \infty} \boldsymbol{P}(\Gamma_n).$$

It follows that the second term on the right side of (4) is equal to $\lim\limits_{n \to \infty} \boldsymbol{P}(\Gamma_n) \boldsymbol{P}(\Delta_i)^{-1}$ which is expressible in terms of the transition matrix. That this limit does not depend on the choice of R is a consequence; in particular since $\{x_t\}$ may be assumed to be well-separable we infer the purely analytical statement that the limit is the same for any sequence R dense in $\boldsymbol{T}$. ▯

The following theorem treats two similar but distinct aspects of taboo probabilities for a post-α process.

Theorem 7. Suppose that the x_+ version is used and that

$$\boldsymbol{P}\{\xi_0(\omega) \neq \infty\} = 1.$$

We have

$$(16) \qquad \boldsymbol{P}\{\xi(u, \omega) \notin H,\ 0 < u < t;\ \xi(t, \omega) = j\} = \sum_{k \notin H} r_k(\Delta; 0)\, {}_Hp_{kj}(t).$$

Furthermore on the set $\{\omega : \alpha(\omega) < t\}$ we haje

$$(17)\qquad \begin{cases} \boldsymbol{P}\{x(u,\omega) \notin H,\ \alpha(\omega) < u < t;\ x(t,\omega) = j \mid \alpha = s\} \\ \qquad\qquad = \sum\limits_{k \notin H} r_k(s, s \mid \Delta)\, {}_H p_{kj}(t-s) \end{cases}$$

for a.a. s in $A(\Delta;\cdot)$ measure.

Proof. Formula (16) is a direct consequence of the strong Markov property, Corollary 3 to Theorem 9.3 and Theorem 6. In contrast, formula (17) is proved as follows. Let $0 < s_0 < t$; then we have, using the method of proof of Theorem 9.2,

$$\boldsymbol{P}\{\alpha(\omega) < s_0;\ x(u,\omega) \notin H,\ \alpha(\omega) < u < t;\ x(t,\omega) = j\}$$

$$\leqq \sum_{m=0}^{n-1} \sum_{k \notin H} \boldsymbol{P}\Big\{\frac{m s_0}{n} \leqq \alpha(\omega) < \frac{(m+1)s_0}{n};\ x\Big(\frac{(m+1)s_0}{n}, \omega\Big) = k;$$

$$x(u,\omega) \notin H,\ \frac{(m+1)s_0}{n} < u < t;\ x(t,\omega) = j\Big\}$$

$$\leqq \int\limits_{(0,s_0)} \sum_{k \notin H} r_k\Big(s, \Big(\Big[\frac{ns}{s_0}\Big] + 1\Big)\frac{s_0}{n} \Big| \Delta\Big)\, {}_H p_{kj}\Big(t - \Big(\Big[\frac{ns}{s_0}\Big] + 1\Big)\frac{s_0}{n}\Big) A(\Delta; ds).$$

The last-written series, say Σ', is dominated by $\sum\limits_k r_k(s, u \mid \Delta)\, p_{kj}(t-u)$ for some u. The latter is uniformly convergent in $u \in [s, t]$ by Theorem 8.4 and DINI's theorem. *A fortiori* Σ' is uniformly convergent in n as well as bounded in s. Hence letting $n \to \infty$ in the above we obtain

$$(18)\qquad \begin{cases} \boldsymbol{P}\{\alpha(\omega) < s_0;\ x(u,\omega) \notin H,\ \alpha(\omega) < u < t;\ x(t,\omega) = j\} \\ \qquad\qquad \leqq \int\limits_{(0,s_0)} \sum\limits_{k \notin H} r_k(s, s \mid \Delta)\, {}_H p_{kj}(t-s)\, A(\Delta; ds). \end{cases}$$

On the other hand, the left member of (18) is clearly the limit, as $\varepsilon \downarrow 0$, of

$$\boldsymbol{P}\{\alpha(\omega) < s_0;\ x(u,\omega) \notin H,\ \alpha(\omega) + \varepsilon < u < t;\ x(t,\omega) = j\}$$

$$\geqq \sum_{m=0}^{n-1} \sum_{k} \boldsymbol{P}\Big\{\frac{m s_0}{n} \leqq \alpha(\omega) < \frac{(m+1)s_0}{n};\ x\Big(\frac{(m+1)s_0}{n}, \omega\Big) = k;$$

$$x(u,\omega) \notin H,\ \frac{(m+2)s_0}{n} < u < t;\ x(t,\omega) = j\Big\}$$

provided that $s_0 n^{-1} < \varepsilon < t - s_0$. Letting first $n \to \infty$ and then $\varepsilon \downarrow 0$ we obtain (18) with "$\leqq$" replaced by "$\geqq$", since the terms with $k \in H$ tend to zero by Theorem 6.2. Thus (18) is true with "$\leqq$" replaced by "$=$" and the truth of this for every s_0 in $(0, t)$ establishes (17). □

The fundamental *first entrance formulas* will now be given, extending (I.9.1). We write ${}_{k,H}p_{ij}$ for ${}_{H_1}p_{ij}$ when $H_1 = H \cup \{k\}$; similarly for other symbols. For an nondecreasing function F on $[0, \infty)$ we write $\int\limits_0^t \ldots dF(s)$ for $\int\limits_{[0,t]} \ldots dF(s)$.

Theorem 8. If $i \neq k$, $k \notin H$ and $t > 0$, we have

$$(19)\qquad {}_Hp_{ij}(t) = {}_{k,H}p_{ij}(t) + \int_0^t {}_Hp_{kj}(t-s)\,d\,{}_HF_{ik}(s)\,;$$

$$(20)\qquad {}_HF_{ij}(t) = {}_{k,H}F_{ij}(t) + \int_0^t {}_HF_{kj}(t-s)\,d\,{}_{j,H}F_{ik}(s)\,.$$

Proof. According to Theorem 6, the validity of the above formulas does not depend on which separable version of the M.C. is used to define the taboo probabilities. We may therefore choose the x_+ version so that

$$x\left({}_H\alpha_{ik}(\omega), \omega\right) = \lim_{t \downarrow \overline{{}_H\alpha_{ik}(\omega)}} x(t, \omega) = k\,,$$

for a.a. ω in ${}_H\Delta_{ik}$. Since ${}_H\alpha_{ik}$ is optional it follows from Theorem 9.5 that for each $l \in \boldsymbol{I}$,

$$(21)\qquad r_l(s, t \mid {}_H\Delta_{ik}) = p_{kl}(t-s)$$

for a.a. s in $A({}_H\Delta_{ik};\cdot)$ measure and $t \geqq s$.

It is clear that

$$\begin{aligned} P_1 &= \boldsymbol{P}\{{}_H\alpha_{ik}(\omega) < t;\ x(u,\omega) \notin H,\ \min[t, \varrho_i(\omega)] < u < t;\ x(t,\omega) = j \mid \Delta_i\} \\ &= \boldsymbol{P}\{{}_H\alpha_{ik}(\omega) < t;\ x(u,\omega) \notin H,\ {}_H\alpha_{ik}(\omega) < u < t;\ x(t,\omega) = j \mid \Delta_i\} \\ &= \int_0^t \boldsymbol{P}\{x(u,\omega) \notin H,\ {}_H\alpha_{ik}(\omega) < u < t;\ x(t,\omega) = j \mid {}_H\alpha_{ik} = s\}\,d\,{}_HF_{ik}(s) \\ &= \int_0^t {}_Hp_{kj}(t-s)\,d\,{}_HF_{ik}(s) \end{aligned}$$

where the last equation follows from (17) and (21). On the other hand, we have

$$\begin{aligned} P_2 = \boldsymbol{P}\{{}_H\alpha_{ik}(t) \geqq t;\ x(u,\omega) \notin H,\ \min[t, \varrho_i(\omega)] < u < t; \\ x(t,\omega) = j \mid \Delta_i\} = {}_{k,H}p_{ij}(t) \end{aligned}$$

by definition. Since ${}_Hp_{ij}(t) = P_1 + P_2$ the formula (19) is proved.

The proof of (20) is similar and will be omitted. ▯

Notes. The content of this section illustrates the caution which is necessary in extending discrete parameter results to the continuous parameter case. Although only a special case of the strong Markov property is involved in the culminating Theorem 8, a "prodigious" amount of preparation, it seems, enters into the proof of this intuitively obvious result. One wonders if the present theory of stochastic process is not still too difficult for applications.

The somewhat contrived definitions given in (3) and (13) have the advantage of yielding some additional formulas such as (12.14) and

(12.15) below. The more natural and convenient definitions would be to replace the "$\min[t, \varrho_t(\omega)]$" in (3) and (13) by "0". The alternatives coincide except when i is stable and $i \in H$.

Theorems 1, 2 and 5 are given in CHUNG [7]. Theorem 3 together with the definition of adjacent set is new. Theorem 4 is new and shows the possibility of a potential-theoretic development; the corollary is due to DOOB [1]. Theorems 6 to 8 are new. Theorem 7 is explicitly stated in that form to illustrate two points: in the proof of (16) a direct resort to the strong Markov property succeeds owing to Corollary 3 to Theorem 9.3 and Theorem 6; while in the proof of (17) such a resort is insufficient even with the help of these, and a re-examination of the proof of the strong Markov property is necessary.

§ 12. Last exit time

The notion of first entrance into a given state after a given time plays a fundamental role in the foregoing developments including the strong Markov property itself; witness the γ_j in (9.6). In the discrete parameter case, there is an obvious dual, namely the last exit from a given state before a given time. This duality is clearly exhibited in the pair of formulas (I.9.1) and (I.9.2). In the continuous parameter case the latter notion is a more intricate one requiring a considerable deepening of probabilistic as well as analytical argument. What lies in the background is the Markov property applied as "time" decreases rather than increases. Specifically, let $\{x_t, t \geqq 0\}$ be a given M.C.; if $\Lambda \in \mathscr{F}_t^0 = \{x_s, 0 \leqq s \leqq t\}$, and $\mathsf{M} \in \mathscr{F}_t' = \mathscr{F}\{x_s, s \geqq t\}$, then we have, as dual to (I.1.4):

$$(1) \qquad \boldsymbol{P}\{\Lambda \mid \mathsf{M};\, x_t(\omega) = i\} = \boldsymbol{P}\{\Lambda \mid x_t(\omega) = i\}.$$

This is seen as follows:

$$\begin{aligned}\boldsymbol{P}\{\Lambda;\, x_t(\omega) = i;\, \mathsf{M}\} &= \boldsymbol{P}\{\Lambda;\, x_t(\omega) = i\}\boldsymbol{P}\{\mathsf{M} \mid x_t(\omega) = i\}\\ &= \boldsymbol{P}\{\Lambda \mid x_t(\omega) = i\}\boldsymbol{P}\{\mathsf{M};\, x_t(\omega) = i\};\end{aligned}$$

cf. (II.9.12) of which the above is a particular case.

For each fixed $t > 0$ we define a stochastic process $\{z_s, 0 \leqq s \leqq t\}$ as follows:

$$z(s, \omega) = x(t - s, \omega).$$

As a consequence of (1), the new process is also a Markov chain but with generally non-stationary transition probabilities. If we assume that $\boldsymbol{P}\{x_0(\omega) = i\} = 1$ and $i \rightsquigarrow j \rightsquigarrow k$, then we have

$$(2) \qquad \boldsymbol{P}\{z(s, \omega) = j \mid z(0, \omega) = k\} = \frac{p_{ij}(t - s)\, p_{jk}(s)}{p_{ik}(t)}.$$

According to Theorem 1.5, the right hand side is positive and finite for $0<s<t$. We shall call $\{z_s, 0 \leqq s \leqq t\}$ the *reverse chain* in $[0, t]$. Its dependence on t and the non-stationarity of its transition probabilities make its direct employment rather difficult. However, certain basic facts may be deduced from our knowledge of the original M.C., such as: if the x_- version is taken then the reverse chain will have almost all its sample functions right lower semicontinuous. The notion of an optional random variable for the reverse chain may be defined and results corresponding to those of the direct M.C. may be gleaned. An important example of this will be given below, in which the random variable τ is clearly "reverse optional" as just mentioned.

Theorem 1. Let τ be a random variable satisfying the following conditions for a fixed $t>0$ and fixed $k \in \boldsymbol{I}$:

(i) $0<\tau(\omega)<t$;
(ii) for each $u>0$: $\{\omega: \tau(\omega)>u\} \in \mathscr{F}\{x_s, s>u\}$;
(iii) $\varliminf_{s \uparrow \tau(\omega)} x(s, \omega)=k$.

For integer $m \geqq 1$ define the r.v. τ_m as follows:

$$\tau_m(\omega)=\frac{n}{2^m} \quad \text{if} \quad \frac{n}{2^m}<\tau(\omega) \leqq \frac{n+1}{2^m}, \quad n \geqq 0;$$

then we have

$$\lim_{m \to \infty} \boldsymbol{P}\{x(\tau_m(\omega), \omega)=k\}=1. \tag{3}$$

Proof. We may suppose that $\boldsymbol{P}\{x_0(\omega)=i\}=1$ for a fixed i. For each $\delta>0$ define

$$\beta_m(\omega)=\max\left\{\frac{n}{2^m}: \delta<\frac{n}{2^m}<\tau(\omega); x\left(\frac{n}{2^m}, \omega\right)=k\right\}.$$

Then $\left\{\omega: \beta_m(\omega)=\frac{n}{2^m}\right\} \in \mathscr{F}\left\{x_u, u \geqq \frac{n}{2^m}\right\}$ by (ii). It follows from (iii) and well-separability that for each m_0 we have

$$\lim_{m \to \infty} \boldsymbol{P}\{\tau_{m_0}(\omega)<\beta_m\}=1. \tag{4}$$

For fixed δ, m_0 and m we put

$$\Lambda_n=\left\{\omega: \delta<\tau_{m_0}(\omega)<\beta_m(\omega)=\frac{n}{2^m}\right\}$$

$$\Lambda_{n,\nu}=\left\{\omega: \delta<\frac{\nu}{2^{m_0}}=\tau_{m_0}(\omega)<\beta_m(\omega)=\frac{n}{2^m}\right\}.$$

Then if $\delta<\frac{\nu}{2^{m_0}}<\frac{n}{2^m}<\frac{\nu+1}{2^{m_0}}$, we have by (ii),

$$\Lambda_{n,\nu}=\left\{\omega: \frac{n}{2^m}=\beta_m(\omega)<\tau(\omega) \leqq \frac{\nu+1}{2^{m_0}}\right\} \in \mathscr{F}\left\{x_u, u \geqq \frac{n}{2^m}\right\}.$$

Hence by the Markov property applied as in (1), we have

$$\boldsymbol{P}\left\{x\left(\frac{\nu}{2^{m_0}}, \omega\right) = k \,|\, \Lambda_{n,\nu}\right\} = \boldsymbol{P}\left\{x\left(\frac{\nu}{2^{m_0}}, \omega\right) = k \,|\, x\left(\frac{n}{2^m}, \omega\right) = k\right\}.$$

The infimum of the last-written probability, for $\frac{n}{2^m} > \delta$ and $0 < \frac{n-\nu}{2^m} < \frac{1}{2^{m_0}}$, is not less than

$$\inf_{\substack{\delta \leqq s \leqq t \\ 0 \leqq h \leqq 2^{-m_0}}} \frac{p_{ik}(s-h)\, p_{kk}(h)}{p_{ik}(s)} = \lambda(m_0),$$

say. Since $p_{ik}(s) > 0$ for $s > 0$ and is continuous in s for $s \geqq 0$, we see that $\lim_{m_0 \to \infty} \lambda(m_0) = 1$. (The invocation of Theorem 1.5 may be avoided by considering only those s for which $p_{ik}(s) > 0$ and using the trivial Theorem 1.4.) Now

$$\boldsymbol{P}\left\{x(\tau_{m_0}(\omega), \omega) = k;\, \tau_{m_0}(\omega) = \frac{\nu}{2^{m_0}} \,|\, \Lambda_n\right\} = \boldsymbol{P}\left\{\tau_{m_0}(\omega) = \frac{\nu}{2^{m_0}} \,|\, \Lambda_n\right\} \times$$
$$\times \boldsymbol{P}\left\{x\left(\frac{\nu}{2^{m_0}}, \omega\right) = k \,|\, \Lambda_{n,\nu}\right\}$$

and consequently summing over ν:

$$\boldsymbol{P}\{x(\tau_{m_0}, \omega) = k \,|\, \Lambda_n\} \geqq \lambda(m_0).$$

Therefore

$$\boldsymbol{P}\{x(\tau_{m_0}(\omega), \omega) = k;\, \delta < \tau_{m_0}(\omega) < \beta_m(\omega)\} \geqq \sum_n \boldsymbol{P}(\Lambda_n)\, \lambda(m_0)$$
$$= \boldsymbol{P}\{\delta < \tau_{m_0}(\omega) < \beta_m(\omega)\}\, \lambda(m_0).$$

Letting $m \to \infty$ and using (4), we have

$$\boldsymbol{P}\{x(\tau_{m_0}(\omega), \omega) = k;\, \delta < \tau_{m_0}(\omega)\} \geqq \boldsymbol{P}\{\delta < \tau_{m_0}(\omega)\}\, \lambda(m_0);$$

letting $m_0 \to \infty$, we have

$$\varliminf_{m_0 \to \infty} \boldsymbol{P}\{x(\tau_{m_0}(\omega), \omega) = k\} \geqq \boldsymbol{P}\{\delta < \tau_{m_0}(\omega)\};$$

letting $\delta \downarrow 0$ we obtain (3). ▯

We proceed to study the last exit times. For an arbitrary $i \in \boldsymbol{I}$ and $t \in \boldsymbol{T}^0$, we define

$$\gamma_i(t, \omega) = \sup\{\overline{S_i(\omega)} \cap [0, t]\}. \tag{5}$$

The separability and measurability of $\{x_t\}$ ensure that $\gamma_i(t) = \gamma_i(t; \cdot)$ is a random variable which will be called the *last exit time from i before time t*. Let $i \neq j$ and $0 \leqq s \leqq t$, and set

$$\Gamma_{ij}(s, t) = \boldsymbol{P}\{\gamma_i(t, \omega) \leqq s;\, x(t, \omega) = j \,|\, x(0, \omega) = i\}. \tag{6}$$

We shall use the notation of § 11, but *redefine* ${}_Hp_{ij}(t)$ as follows:

$$(7)\qquad {}_Hp_{ij}(t)=\boldsymbol{P}\{x(s,\omega)\notin H,\ 0<s<t;\ x(t,\omega)=j\,|\,x(0,\omega)=i\}.$$

Thus ${}_Hp_{ij}(t)\equiv 0$ if $i\in H$ or $j\in H$ by separability. The new definition coincides with the old one in (11.3) except when i is stable and $i\in H$, and yields the semigroup property

$$(8)\qquad {}_Hp_{ij}(s+t)=\sum_k {}_Hp_{ik}(s)\,{}_Hp_{kj}(t)$$

without exception; in fact we may regard $({}_Hp_{ij}(\cdot))$ as a substochastic transition matrix indexed by $(\boldsymbol{I}-H)\times(\boldsymbol{I}-H)$. We shall need it below only in the case $H=\{i\}$; note then for each $k\neq i$:

$$(9)\qquad \sum_j {}_ip_{kj}(t)=1-F_{ki}(t).$$

It follows also from the definitions that for $0\leqq s<t$:

$$(10)\qquad \Gamma_{ij}(s,t)=\sum_k p_{ik}(s)\,{}_ip_{kj}(t-s),$$

where the summation is really over $\boldsymbol{I}-\{i\}$. We set for $0\leqq s\leqq t$:

$$(11)\qquad \Gamma_i(s,t)=\boldsymbol{P}\{\gamma_i(t,\omega)\leqq s\,|\,x(0,\omega)=i\},$$

then we have for $0\leqq s<t$:

$$\Gamma_i(s,t)=\sum_{j\neq i}\Gamma_{ij}(s,t)=\sum_{k\neq i}p_{ik}(s)\left[1-F_{ki}(t-s)\right].$$

For $s=t$ this relation is no longer valid, since the first term above equals one by (11) while the last equals $1-p_{ii}(t)$. It follows that the distribution of $\gamma_i(t)$ is continuous in $[0,t)$ but has a jump of magnitude $p_{ii}(t)$ at t.

The basic decomposition by means of the last exit time is given in the next theorem.

Theorem 2. For $i\neq j$ and $t\geqq 0$ we have:

$$(12)\qquad p_{ij}(t)=\lim_{m\to\infty}\sum_{\nu=1}^{[2^m t]}p_{ii}\left(\frac{\nu-1}{2^m}\right)\sum_{k\neq i}p_{ik}\left(\frac{1}{2^m}\right)\,{}_ip_{kj}\left(t-\frac{\nu}{2^m}\right).$$

Furthermore for each m, the inequality "$\geqq$" holds in (12) without the limit.

Proof. The exhibited term in the first sum above represents the probability of the set

$$\begin{aligned}A_\nu&=\left\{\omega:\ x\left(\frac{\nu-1}{2^m},\omega\right)=i,\ x(s,\omega)\neq i,\ \frac{\nu}{2^m}\leqq s<t;\ x(t,\omega)=j\right\}\\&=\left\{\omega:\ x\left(\frac{\nu-1}{2^m},\omega\right)=i;\ \frac{\nu-1}{2^m}<\gamma_i(t,\omega)\leqq\frac{\nu}{2^m};\ x(t,\omega)=j\right\}.\end{aligned}$$

The asserted inequality follows at once by summing over ν. Now define $\gamma_i^{(m)}(t)$ in terms of $\gamma_i(t)$ as τ_m is defined in terms of τ in Theorem 1; and put

$$B = \bigcup_{\nu=1}^{[2^m t]} \left\{\omega : x\left(\frac{\nu-1}{2^m}, \omega\right) \neq i;\ \frac{\nu-1}{2^m} < \gamma_i(t, \omega) \leqq \frac{\nu}{2^m};\ x(t, \omega) = j\right\}$$

$$= \left\{\omega : x(\gamma_i^{(m)}(t, \omega), \omega) \neq i;\ \gamma_i(t, \omega) \leqq \frac{[2^m t]}{2^m};\ x(t, \omega) = j\right\}.$$

Then we have clearly

$$(13) \qquad p_{ij}(t) = \sum_{\nu=1}^{[2^m t]} \boldsymbol{P}(A_\nu) + \boldsymbol{P}(B) + \boldsymbol{P}\left\{\frac{[2^m t]}{2^m} < \gamma_i(t, \omega)\right\}.$$

On the set $\Delta = \{\omega : x(0, \omega) = i,\ x(t, \omega) = j\}$, we have $0 < \gamma_i(t, \omega) < t$ a.e. by Theorem 6.2 applied to i at 0 and to j at t. Next for each $u > 0$,

$$\{\omega : \gamma_i(t, \omega) > u\} = \{\omega : S_i(\omega) \cap (u, t) \neq 0\}$$

so that $\gamma_i(t)$ satisfies condition (ii) of Theorem 1 for τ. Condition (iii) there is also satisfied by the definition of $\gamma_i(t)$ and Theorem 6.1 (recall that $\boldsymbol{I}$ has been taken to be the set of positive integers in § 11.7). We may therefore apply Theorem 1 above to the set Δ and the random variable $\gamma_i(t)$, either by taking the conditional probability $\boldsymbol{P}(\cdot \mid \Delta)$ or perhaps more directly by taking the intersection of every set in Theorem 1 and its proof with Δ. We conclude that $\lim_{m \to \infty} \boldsymbol{P}(B) = 0$. Since the last term in (13) also converges to zero as $m \to \infty$, (12) follows from (13) by letting $m \to \infty$. ▯

We are going to transform the right hand side of (12) into a convolution integral. First we shall treat the second sum in (12). For $i \neq j$, $\delta > 0$ and $s \geqq 0$ let us put

$$(14) \qquad \varphi_{ij}(\delta; s) = \frac{1}{\delta} \sum_k p_{ik}(\delta)\, {}_i p_{kj}(s);$$

note that the summation above is really over $k \neq i$ by the new definition of ${}_i p_{kj}$. For each $s \geqq 0$, $\varphi_{ij}(\delta; s)$ is continuous in $\delta > 0$ and for each $\delta > 0$, $\varphi_{ij}(\delta; s)$ is continuous in $s \geqq 0$. Both these assertions follow from the uniform convergence of $\sum_k p_{ik}(\delta)$ in every finite interval and the continuity of the ${}_i p_{kj}$'s.

Theorem 3. As $\delta \downarrow 0$, $\varphi_{ij}(\delta; s)$ converges to a function g_{ij} which is measurable and bounded everywhere; and we have

$$(15) \qquad p_{ij}(t) = \int_0^t p_{ii}(t - s)\, g_{ij}(s)\, ds.$$

Proof. Consider the discrete skeleton $\{x_{n\delta},\ n\geqq 0\}$ for a fixed $\delta>0$, and let ${}_ip_{kj}^{(n)}(\delta)$ be the taboo probability in this d.p. M.C., defined in §13 below; clearly ${}_ip_{kj}^{(n)}(s)\leqq {}_ip_{kj}(n\delta)$, the latter being defined in (7). We have as a special case of (I.9.2),

$$p_{ik}(m\delta)=\sum_{\nu=0}^{m-1} p_{ii}(\nu\delta)\ {}_ip_{ik}^{(m-\nu)}(\delta).$$

It follows from the preceding remark that

$${}_ip_{ik}^{(m-\nu)}(\delta)=\sum_{l\neq i} p_{il}(\delta)\ {}_ip_{lk}^{(m-\nu-1)}(\delta)\geqq \sum_{l} p_{il}(\delta)\ {}_ip_{lk}\big((m-\nu-1)\delta\big).$$

Hence we have by (8) with $H=\{i\}$:

$$\begin{aligned}\varphi_{ij}(m\delta;s)&\geqq \frac{1}{m\delta}\sum_{\nu=0}^{m-1} p_{ii}(\nu\delta)\sum_{k}\sum_{l} p_{il}(\delta)\,{}_ip_{lk}\big((m-\nu-1)\delta\big)\,{}_ip_{kj}(s)\\ &=\frac{1}{m\delta}\sum_{\nu=0}^{m-1} p_{ii}(\nu\delta)\sum_{l} p_{il}(\delta)\,{}_ip_{lj}\big((m-\nu-1)\delta+s\big)\\ &\geqq \frac{1}{m\delta}\sum_{\nu=0}^{m-1} p_{ii}(\nu\delta)\sum_{l} p_{il}(\delta)\,{}_ip_{lj}(s)\,{}_ip_{jj}\big((m-\nu-1)\delta\big).\end{aligned}$$

Given $\varepsilon>0$, there exists $\eta=\eta\,(i,j,\varepsilon)>0$ such that if $m\delta\leqq\eta$ then

$$\min_{0\leqq\nu\leqq m} p_{ii}(\nu\delta)\min_{0\leqq\nu\leqq m} {}_ip_{jj}\big((m-\nu-1)\delta\big)\geqq 1-\varepsilon,$$

since $\lim\limits_{t\downarrow 0}\ {}_ip_{jj}(t)=1$. It follows that

$$(16)\qquad \varphi_{ij}(m\delta;s)\geqq\frac{1}{m\delta}\,m(1-\varepsilon)\sum_{l} p_{il}(\delta)\,{}_ip_{lj}(s)=(1-\varepsilon)\,\varphi_{ij}(\delta;s).$$

Taking $\varepsilon=\frac{1}{2}$, $\eta\,(i,j,\frac{1}{2})=\eta_0$ and $m=[\eta_0/\delta]$, we have

$$(17)\qquad \sup_{0\leqq\delta\leqq\eta_0/2;\ s\geqq 0}\varphi_{ij}(\delta;s)\leqq 2\varphi_{ij}(m\delta;s)\leqq\frac{2}{m\delta}\leqq\frac{2}{\eta_0-\delta}\leqq\frac{4}{\eta_0}.$$

Next, define for $s\geqq 0$:

$$(18)\qquad g_{ij}(s)=\overline{\lim_{s\downarrow 0}}\ \varphi_{ij}(\delta;s).$$

Letting $\delta\uparrow 0$ and $m\delta\to t$, where $0<t<\eta(\varepsilon)$, in (16), we obtain

$$\varphi_{ij}(t;s)\geqq(1-\varepsilon)\,g_{ij}(s),$$

since $\varphi_{ij}(\cdot,s)$ is continuous for fixed s. Thus

$$\underline{\lim_{t\downarrow 0}}\ \varphi_{ij}(t;s)\geqq(1-\varepsilon)\ \overline{\lim_{s\downarrow 0}}\,\varphi_{ij}(\delta;s),$$

and consequently (18) is true with $\overline{\lim}$ replaced by lim. Together with (17) this proves the first sentence of the theorem. Note that the argument above sharpens one in the proof of Theorem 2.5, which becomes indeed the particular case for $s=0$. ▯

Next, introducing the jump function

$$\pi_i(\delta;s)=\sum_{\nu=1}^{[s/\delta]} p_{ii}((\nu-1)\delta)\,\delta;$$

and with $\delta=2^{-m}$, we rewrite the assertions of Theorem 2 as follows:

$$(19_1)\qquad p_{ij}(t)=\lim_{\delta\downarrow 0}\int_0^t \varphi_{ij}(\delta;t-s)\,d\pi_i(\delta;s);$$

$$(19_2)\qquad p_{ij}(t)\geqq\int_0^t \varphi_{ij}(\delta;t-s)\,d\pi_i(\delta;s).$$

Introducing also the integrals

$$P_{ij}(t)=\int_0^t p_{ij}(s)\,ds,\qquad \Phi_{ij}(\delta;t)=\int_0^t \varphi_{ij}(\delta;s)\,ds,\qquad G_{ij}(t)=\int_0^t g_{ij}(s)\,ds;$$

we have from (19_2)

$$(20)\qquad P_{ij}(t)\geqq\int_0^t \Phi_{ij}(\delta;t-s)\,d\pi_i(\delta;s).$$

Owing to this we may integrate (19_1) and use dominated convergence to conclude

$$(21)\qquad P_{ij}(t)=\lim_{\delta\downarrow 0}\int_0^t \Phi_{ij}(\delta;t-s)\,d\pi_i(\delta;s).$$

It follows from what we have proved above that

$$(22)\qquad G_{ij}(t)=\lim_{\delta\downarrow 0}\Phi_{ij}(\delta;t);$$

and furthermore it follows from (16) that given any $\varepsilon>0$, there exists $\delta_0(\varepsilon)>0$ such that

$$(23)\qquad \Phi_{ij}(\delta_1;s)\geqq(1-\varepsilon)\,\Phi_{ij}(\delta_2;s)\quad\text{for}\quad 0<\delta_2<\delta_1<\delta_0(\varepsilon).$$

Thus the sequence of nonnegative continuous functions $\Phi_{ij}(\delta,\cdot)$ converges to the continuous function $G_{ij}(\cdot)$ with the "quasi-monotonicity" indicated in (23). A trivial modification of DINI's theorem on monotone convergence (see e.g. TITCHMARSH [1; p. 13]) implies that the convergence in (22) must be uniform in every finite interval. On the other hand, we have

$$\lim_{\delta\downarrow 0}\pi_i(\delta;s)=\int_0^s p_{ii}(u)\,du$$

since p_{ii} is continuous. The last two assertions allow us to pass to the limit in both integrator and integrand in (21) and obtain

$$P_{ij}(t) = \int_0^t G_{ij}(t-s)\, p_{ii}(s)\, ds, \tag{24}$$

the integrated form of (15). Now a simple transformation shows that the right hand side in (24) is equal to

$$\begin{aligned}\int_0^t \left[\int_0^{t-s} g_{ij}(u)\, du\right] p_{ii}(s)\, ds &= \int_0^t p_{ii}(s)\, ds \int_s^t g_{ij}(u-s)\, du \\ &= \int_0^t du \int_0^u g_{ij}(u-s)\, p_{ii}(s)\, ds.\end{aligned}$$

Since P_{ij} is also the integral of p_{ij}, it follows that (15) must hold for almost every t. But the right member as well as the left is a continuous function of t by the continuity of p_{ii} and boundedness of g_{ij}; hence (15) holds for every t as asserted. □

For further analysis of the function g_{ij} we need the following purely analytical lemma which is also useful elsewhere; it may be regarded as a companion to Theorem 2.3, progressing from continuity to continuous differentiability.

Theorem 4. Let (p_{ij}), $i, j \in \boldsymbol{I}$, be a matrix of finite functions on $\boldsymbol{T}^0$ satisfying the conditions (A), (C) of § II.1 and (E) of § II.2 [condition (B) being omitted]. Let $\{G_j, j \in \boldsymbol{I}\}$ be nonegative, nondecreasing finite functions on $\boldsymbol{T}$ satisfying the equations, for $s, t \in \boldsymbol{T}$:

$$G_j(s+t) - G_j(t) = \sum_i G_i(s)\, p_{ij}(t), \qquad j \in \boldsymbol{I}; \tag{25}$$

$$[\text{or } G_i(t+s) - G_i(t) = \sum_j p_{ij}(t)\, G_j(s), \qquad i \in \boldsymbol{I}]. \tag{25'}$$

Then each G_j has a continuous derivative g_j in $\boldsymbol{T}$ satisfying for $s \in \boldsymbol{T}^0$, $t \in \boldsymbol{T}$:

$$G_j'(s+t) = \sum_i G_i'(s)\, p_{ij}(t); \tag{26}$$

$$[\text{or } G_i'(t+s) = \sum_j p_{ij}(t)\, G_j'(s).] \tag{26'}$$

Proof. Taking difference quotients with respect to s in (25), passing to the limit and using Fatou's lemma, we obtain

$$G_j'(s+t) \geqq \sum_i G_i'(s)\, p_{ij}(t) \tag{27}$$

where G_i' denotes the right-hand lower derivate of G_i. On the other hand, since each G_i is clearly nondecreasing, Fubini's theorem on differentiation cited in the proof of Theorem 3.1 asserts: for each t there exists a set $Z_0(t)$ of measure zero, such that if $s \notin Z_0(t)$ then each G_i has a finite nonnegative ordinary derivative satisfying (27) with equality. Each such derivative must of course coincide with the G'_i above. By FUBINI's theorem on product measures, there is then a set Z of measure zero and for each $s \notin Z$ a set $Z(s)$ of measure zero such that

$$G'_j(s+t) = \sum_i G'_i(s) p_{ij}(t), \quad \text{if} \quad s \notin Z,\ t \notin Z(s). \tag{28}$$

Now let $s_0 \notin Z$ and suppose that for a certain t_0, (28) is false so that in view of (27) we have the strict inequality

$$G'_j(s_0+t_0) > \sum_i G'_i(s_0) p_{ij}(t_0). \tag{29}$$

Then for every $t > t_0$ it follows from (27) and (29) that

$$\begin{aligned} G'_j(s_0+t) &\geqq \sum_i G'_i(s_0+t_0) p_{ij}(t-t_0) \\ &\geqq \sum_i \Big[\sum_k G'_k(s_0) p_{ki}(t_0)\Big] p_{ij}(t-t_0) = \sum_k G'_k(s_0) p_{kj}(t). \end{aligned}$$

This contradicts (28) and consequently (28) must hold for every t, provided $s \notin Z$. Finally for an arbitrary $s > 0$ we may write $s = s_0 + s_1$ where $s_0 \notin Z$, and then it follows from what has just been shown that

$$\begin{aligned} G'_j(s+t) = G'_j(s_0+s_1+t) &= \sum_i G'_i(s_0) p_{ij}(s_1+t) = \sum_i G'_i(s_0) \sum_k p_{ik}(s_1) p_{kj}(t) \\ &= \sum_k \Big[\sum_i G'_i(s_0) p_{ik}(s_1)\Big] p_{kj}(t) = \sum_k G_k(s_0+s_1) p_{kj}(t) \\ &= \sum_k G_k(s) p_{kj}(t). \end{aligned}$$

Hence (28) holds for every $s \in \boldsymbol{T}^0$ and $t \in \boldsymbol{T}^0$. It now follows from Theorem 2.3 that each G'_i may be extended to be continuous in $\boldsymbol{T}$ and the theorem is proved. ▯

Before we return to the previous discussion let us first complete the first entrance formula as given in (11.19) with H empty and $j = k$:

$$p_{ik}(t) = \int_0^t p_{kk}(t-s)\, dF_{ik}(s), \qquad i \neq k.$$

Theorem 5. For $i \neq k$, F_{ik} has a continuous derivative f_{ik} satisfying

$$\sum_j {}_k p_{ij}(t) f_{jk}(s) = f_{ik}(t+s), \qquad t \geqq 0,\ s > 0; \tag{30}$$

and

$$p_{ik}(t) = \int_0^t f_{ik}(s) p_{kk}(t-s)\, ds. \tag{31}$$

Proof. It is immediate from the meaning of the probabilities involved that for i, $k \in \boldsymbol{I}$ and $t \geqq 0$, $s > 0$:

$$\sum_j {}_k p_{ij}(t) F_{jk}(s) = F_{ik}(t+s) - F_{ik}(t). \tag{32}$$

For fixed i and k we apply Theorem 4 in the form of (25′) and (26′) with ${}_k p_{ij}$ for p_{ij} and F_{jk} for G_j; the conclusions follow. ☐

Theorem 6. The convergence of $\varphi_{ij}(\delta; s)$ to $g_{ij}(s)$ as $\delta \downarrow 0$ is uniform in every finite interval of s; g_{ij} is continuous in $\boldsymbol{T}$; and we have for every i and k:

$$\sum_j g_{ij}(s)\, {}_i p_{jk}(t) = g_{ik}(s+t), \qquad s > 0,\ t \geqq 0. \tag{33}$$

Proof. It follows from (8) and (14) that for each $\delta > 0$:

$$\sum_j \varphi_{ij}(\delta; s)\, {}_i p_{jk}(t) = \varphi_{ij}(\delta; s+t)$$

and consequently by integration:

$$\sum_j \Phi_{ij}(\delta; s)\, {}_i p_{jk}(t) = \Phi_{ik}(\delta; s+t) - \Phi_{ik}(\delta; t). \tag{34}$$

From (20) we have if $0 < s < t$:

$$P_{ij}(t) \geqq \Phi_{ij}(\delta; s) \int_{t-s}^{t} d\pi_i(\delta; s)$$

where the last-written integral is bounded away from zero independently of δ. Hence the series $\sum_j \Phi_{ij}(\delta; s)$ converges uniformly in δ for each fixed s, and letting $\delta \downarrow 0$ in (34) we obtain

$$\sum_j G_{ij}(s)\, {}_i p_{jk}(t) = G_{ik}(s+t) - G_{ik}(t), \tag{35}$$

which is the exact dual to (32). Applying Theorem 4 we conclude that each G_{ij} has a continuous derivative G'_{ij} satisfying

$$\sum_j G'_{ij}(s)\, {}_i p_{jk}(t) = G'_{ik}(s+t).$$

Since the argument under (24) leads to

$$p_{ij}(t) = \int_0^t p_{ii}(t-s)\, G'_{ij}(s)\, ds,$$

we have, substituting back into (14),

$$\varphi_{ij}(\delta; s) = \frac{1}{\delta} \sum_k \int_0^\delta p_{ii}(\delta - u)\, G'_{ik}(u)\, {}_i p_{kj}(s)\, du$$

$$= \frac{1}{\delta} \int_0^\delta p_{ii}(\delta - u)\, G'_{ij}(u+s)\, ds.$$

Letting $\delta\downarrow 0$ and using the continuity of G'_{ij}, we achieve our goal of identification:

$$g_{ij}(s)=\lim_{\delta\downarrow 0}\varphi_{ij}(\delta;s)=G'_{ij}(s).$$

Now that g_{ij} is known to be continuous, the uniform convergence of $\varphi_{ij}(\delta;\cdot)$ to it follows in the same way as that of $\Phi_{ij}(\delta;\cdot)$ to G_{ij} in the proof of Theorem 3. ▯

Corollary. For $s>0$,

$$g_i(s)=\sum_{j\neq i}g_{ij}(s) \tag{36}$$

is continuous and nonincreasing in s. We have, recalling (3.1):

$$g_{ij}(0)=p'_{ij}(0)=q_{ij}; \tag{37}$$

$$g_i(0)\ =g_i(0+)=-p'_{ii}(0)=q_i. \tag{38}$$

Proof. Summing (15) over $j\neq i$ we see that $g_i(s)<\infty$ for a.e. s. Summing (33) over $k\neq i$ we obtain

$$g_i(s+t)=\sum_{j\neq i}g_{ij}(s)\,[1-F_{ji}(t)];$$

hence $g_i(s)<\infty$ for every $s>0$ and is nonincreasing. It is also continuous since each F_{ji} is. Since g_{ij} is continuous at 0, (37) follows at once from (15); while (38) follows similarly upon first summing (15) over $j\neq i$. ▯

We now return to the last exit time $\gamma_i(t)$ and state its fundamental properties.

Theorem 7. For each t, both $\Gamma_{ij}(s,t)$ and $\Gamma_i(t)$ as defined in (6) and (11) respectively have continuous derivatives with respect to s given by

$$\left.\begin{aligned}&\frac{\partial}{\partial s}\Gamma_{ij}(s,t)=p_{ii}(s)g_{ij}(t-s), &(39)\\ &\frac{\partial}{\partial s}\Gamma_i(s,t)=p_{ii}(s)g_i(t-s), &(40)\end{aligned}\right\}\quad 0\leqq s<t.$$

We have, as one version of the conditional probability below:

$$\boldsymbol{P}\{x(t,\omega)=j\,|\,x(0,\omega)=i;\gamma_i(t,\omega)=s\}=\frac{g_{ij}(t-s)}{g_i(t-s)}. \tag{41}$$

Proof. We have by substituting (15) into (10) and using (8):

$$\Gamma_{ij}(s,t)=\sum_k\int_0^s p_{ii}(u)g_{ik}(s-u)\,{}_ip_{kj}(t-s)\,du=\int_0^s p_{ii}(u)\,g_{ij}(t-u)\,du.$$

Summing over $j\neq i$ we deduce

$$\Gamma_i(s,t)=\int_0^s p_{ii}(u)g_i(t-u)\,du.$$

These two equations imply (39) and (40). Now we can rewrite (15) as

$$p_{ij}(t) = \int_0^t \frac{g_{ij}(t-s)}{g_i(t-s)} \, d_s \Gamma_i(s, t)$$

which implies (41). ▯

As an important application of Theorem 1, we are going to establish the continuous differentiability of each p_{ij} in $\boldsymbol{T}^0$, together with the convergence properties of the series obtained by termwise differentiation of (1.B) and (1.C). These results will extend the results of § 3 where either i or j is assumed to be stable. The proof will be divided into several steps.

1°. There exists $a_i > 0$ such that p_{ii} is absolutely continuous in $[0, a_i]$.

Proof.

$$(42)\quad 1 - p_{ii}(t) = \int_0^t p_{ii}(t-s) g_i(s)\, ds = G_i(t) - \int_0^t [1 - p_{ii}(t-s)]\, dG_i(s)$$

where $G_i(t) = \int_0^t g_i(s)\, ds < \infty$. Let a_i be chosen such that $G_i(a_i) < 1$. Define $g_i^{(1)} = g_i$ and for $n \geqq 2$,

$$g_i^{(n)}(t) = \int_0^t g_i^{(n-1)}(t-s) g_i^{(1)}(s) ds\ ,$$

$$G_i^{(n)}(t) = \int_0^t g_i^{(n)}(s)\, ds.$$

If $0 \leqq t \leqq a_i$, a simple induction shows that $G_i^{(n)}(t) \leqq [G_i(t)]^n$. Hence the series

$$(43)\qquad H_i(t) = \sum_{n=1}^{\infty} (-1)^{n-1} G_i^{(n)}(t)$$

converges absolutely and satisfies the following equation in $[0, a_i]$:

$$(44)\qquad \int_0^t H_i(t-s)\, dG_i(s) = G_i(t) - H_i(t).$$

Regarding (44) as an integral equation in the unknown H_i for given G_i, we see that there cannot be two distinct bounded solutions by considering the supremum of their difference. Since H_i is bounded, it follows that $1 - p_{ii}(t) \equiv H_i(t)$ in $[0, a_i]$. Since H_i is the sum of an absolutely convergent series of absolutely continuous functions it is itself absolutely continuous in $[0, a_i]$.

For $0 \leqq s < t < \infty$ let $v_{ij}(s, t)$ denote the *total variation* of p_{ij} in $[s, t]$. From 1° it follows by a known theorem that if $0 \leqq u \leqq a_i$ and p'_{ii} denotes a derivative of p_{ii} there, we have

$$v_{ii}(0, u) = \int_0^u |p'_{ii}(t)|\, dt < \infty .$$

2°. For every i and j, there exist functions $\alpha_i(u)$ and $\beta_j(u)$ which are bounded in every finite interval such that for all $u \geqq 0$ we have

$$v_{ij}(0, u) \leqq \alpha_i(u) p_{ij}(u), \tag{45}$$

$$v_{ij}(0, u) \leqq p_{ij}(u) \beta_j(u). \tag{46}$$

Proof. Let $i \neq j$; it follows from 1° and (15) that p_{ij} is absolutely continuous in $[0, a_i]$, and furthermore we have there

$$\int_0^u |p'_{ij}(t)|\, dt \leqq \int_0^u \left[g_{ij}(t) + \int_0^t g_{ij}(t-s)\, |p'_{ii}(s)|\, ds \right] dt \leqq G_{ij}(u)\,[1 + v_{ii}(0, a_i)] .$$

It follows also from (15) that if $0 \leqq u \leqq a_i$ then

$$G_{ij}(u) \leqq \frac{p_{ij}(u)}{m_i(a_i)}$$

where $m_i(t) = \min\limits_{0 \leqq s \leqq t} p_{ii}(s)$. Combining the last two inequalities we see that (45) is true if $i \neq j$ and $0 \leqq u \leqq a_i$, with $\alpha_i(u)$ taken to be some constant α_i. If $i = j$ and $0 \leqq u \leqq a_i$, (45) is trivially true by 1° and a suitable choice of α_i. Now we have clearly by (1.C),

$$v_{ij}(s, s+u) = \sum_k v_{ik}(0, u) p_{kj}(s);$$

hence the general case of (45) is obtained by adding the total variation over successive intervals. The proof of (46) is dual.

3°. For every i and j, p_{ij} is absolutely continuous in every finite interval.

Proof. It follows from (45) that

$$\sum_k v_{ik}(0, a_i) p_{kj}(t) < \infty \tag{47}$$

for every t. Hence we have if $0 \leqq s \leqq a_i$, using the absolute continuity of all p_{ik} in $[0, a_i]$ proved in 2°:

$$p_{ij}(s+t) - p_{ij}(t) = \sum_k \int_0^s p'_{ik}(u) p_{kj}(t)\, du = \int_0^s \sum_k p'_{ik}(u) p_{kj}(t)\, du ,$$

where the last series converges absolutely for a.a. u in $[0, a_i]$ and all t, by (47). This equation shows that for each fixed t, p_{ij} is absolutely continuous in $[t, t+a_i]$ and consequently it is so in every finite interval.

4°. For every i and j, p_{ij} has a continuous derivative in $(0, \infty)$.

Proof. Let $M_{ij}(t) = \max_{0 \leqq s \leqq t} p_{ij}(s)$; then

$$M_{ij}(t) \leqq v_{ij}(0, t).$$

It follows from (45) and (46) that for every s and t,

$$\sum_k v_{ik}(0, s)\, v_{kj}(0, t) < \infty;$$

a fortiori,

$$\sum_k M_{ik}(s) \int_0^t |p'_{kj}(u)|\, du < \infty. \tag{48}$$

Hence for each s and a.a. t (depending on s) we have

$$\sum_k M_{ik}(s)\, |p'_{kj}(t)| < \infty. \tag{49}$$

By Fubini's theorem, (49) is also valid if $t \notin Z$ and $s \notin Z_t$ where Z and Z_t are sets of measure zero. But $M_{ik}(s)$ is nondecreasing in s, hence (49) is valid for $t \notin Z$ and all s. In particular if $t \notin Z$, the series $\sum_k p_{ik}(s)\, p'_{kj}(t)$ converges (absolutely and) uniformly for s in every finite interval and so is a continuous function of s. Next, for each s we have

$$p_{ij}(s+t) - p_{ij}(s) = \sum_k p_{ik}(s) \int_0^t p'_{kj}(u)\, du = \int_0^t \sum_k p_{ik}(s)\, p'_{kj}(u)\, du$$

where the second equation is justified by (48); consequently we have for a.a. t (depending on s),

$$p'_{ij}(s+t) = \sum_k p_{ik}(s)\, p'_{kj}(t). \tag{50}$$

By Fubini's theorem, (50) is also valid if $t \notin Z$ and $s \notin Z_t$ where Z and Z_t may be taken to be the same sets as before. For a fixed $t > 0$, $t \notin Z$, the left member of (50), which is a derivative of the absolutely continuous function $p_{ij}(s+t)$ of s, coincides for a.a. s with the right member which is a continuous function of s. Hence $p_{ij}(s+t)$ has a continuous derivative with respect to s. This being true for arbitrarily small values of t we conclude that p_{ij} has a continuous derivative in $(0, \infty)$ satisfying (50). A dual argument shows that it also satisfies the dual equation

$$p'_{ij}(t+s) = \sum_k p'_{ik}(t)\, p_{kj}(s). \tag{51}$$

We state the results as follows.

Theorem 8. Let (p_{ij}) be a standard transition matrix. Then every element p_{ij} has a continuous derivative in $(0, \infty)$ satisfying (50) and (51) for all $s > 0$ and $t > 0$, where the series are absolutely convergent.

We have also for each i,

$$\sum_k |p'_{ik}(t)| < \infty, \quad \sum_k p'_{ik}(t) = 0, \quad 0 < t < \infty. \tag{52}$$

If $i \neq j$, p'_{ij} is also continuous at zero.

Proof. To show e.g. the absolute convergence of (50), we prove the stronger result that (49) holds for all $s > 0$ and $t > 0$. This has been proved in 4° if $t \notin Z$. For an arbitrary t this follows by choosing $t_0 < t$, $t_0 \notin Z$ and observing that

$$\begin{aligned}\sum_k M_{ik}(s)|p'_{kj}(t)| &\leqq \sum_k M_{ik}(s) \sum_l p_{kl}(t-t_0)|p'_{lj}(t_0)| \\ &\leqq \sum_l M_{il}(s+t-t_0)|p'_{lj}(t_0)| < \infty.\end{aligned}$$

Next, it follows from (51) that the first sum in (52) is nonincreasing in t, and consequently from (45) that it is finite for all $t > 0$. Because of this, we can sum (51) over all j and invert the repeated sum on the right side. It follows that the second sum in (52) is a constant for all $t > 0$. This constant must be zero since we may integrate it from 0 to u termwise by dominated convergence and the reuslt is equal to zero for all $u > 0$. Only the last assertion of the theorem remains to be proved. We have from (15),

$$p'_{ij}(t) = g_{ij}(t) + \int_0^t g_{ij}(t-s)\, p'_{ii}(s)\, ds;$$

consequently $\lim_{t \downarrow 0} p'_{ij}(t) = g_{ij}(0) = p'_{ij}(0)$ by 3° and Corollary to Theorem 6. ▯

Corollary. For every i and j, we have

$$\lim_{t \to \infty} p'_{ij}(t) = 0. \tag{53}$$

Proof. By (51), (52) and Theorem 10.1, we have for any $s > 0$:

$$\lim_{t \to \infty} p'_{ij}(t) = \sum_k p'_{ik}(s)\, \pi_{kj}.$$

Thus the limit in (53) exists. It must then be zero since $\int_a^b p'_{ij}(t)\, dt = p_{ij}(b) - p_{ij}(a)$ is bounded as a and b tend to infinity. ▯

Theorem 8 leaves open the question of continuity of $p'_{ii}(t)$ at $t = 0$. For a stable i we know that there is continuity from Theorem 3.1, or by differentiating the first equation in (42) and using (38) and 1° above. For an instantaneous i the situation offers a radical contrast as follows: we have always

$$\lim_{t \downarrow 0} p'_{ii}(t) = -\infty, \tag{54}$$

as an immediate consequence of (2.6), Theorem 8 and the mean value theorem of differential calculus; but it is possible that

$$\overline{\lim_{t\downarrow 0}}\, p'_{ii}(t) = +\infty. \tag{55}$$

Notes. This section is an amplification of the Addenda in the first edition of the book. Theorem 1 has been a missing link there which will be needed again in § 16 below. A particular case of this, which would have been adequate for our purposes and corresponds to the case where τ is the first entrance time into k in the reverse chain in $[0, t]$, has been supplied to the Russian translation of the first edition. The last exit representation formula (15) was first proved by JURKAT [3], though in the case of a stable i this was anticipated by (15.27). His method is analytic, does not use (1.B) and relies on applying HELLY's theorem to obtain a sequence of $\delta\downarrow 0$ for which $\sum\limits_{\nu\delta\leqq t} {}_ip_{ik}^{(\nu)}(\delta)$ converges to some function $G_{ik}(t)$. The proof given here is taken from CHUNG [8]. The idea of considering $\gamma_i(t)$ as a first entrance time in the reverse chain just mentioned leads to another probabilistic proof. For details of this and an interesting special case where a simple "duality principle" applies, see CHUNG [8] and [9]. Theorems 4 and 5 also appear in the last cited paper; the former was proved independently by NEVEU [2]. Theorem 7 becomes more interesting when compared with the analogues in boundary theory; see CHUNG [12] and [14]. The major part of Theorem 8 is due to D. ORNSTEIN [1] whose proof is more of the "discrete" type. Some auxiliary results were obtained by the author but the proof given here is a modified version of the one given in JURKAT [3]. The last paper contains also a number of interesting results not covered here. The corollary to Theorem 8 was inadvertently overlooked in the first edition, as pointed out by REUTER; if $\boldsymbol{I}$ forms a recurrent class the stronger result $\lim\limits_{t\to\infty}\sum\limits_j |p'_{ij}(t)| = 0$ has been proved by OREY [3]. The question about the continuity of $p'_{ii}(t)$ at $t=0$ for an instantaneous i was raised at the end of the first edition and answered by GERALD J. SMITH [1] who constructed an example in which (55) holds, by an ingenious "perturbation" of the probabilistic analysis of Example 3 in § 20 below. According to him (oral communication) every "number" in $[-\infty, +\infty]$ is a possible values of $\lim\limits_{t\downarrow 0} p'_{ii}(t)$ for an instantaneous i. For a review of other analytic problems concerning the elements of a transition matrix see KENDALL [5].

§ 13. Ratio limit theorems; discrete approximations

In this section we apply the Laplace transform to some instances of the first entrance formulas (Theorem 11.8) to obtain analytical results which extend those of §§ I.9—10. Let p and F stand respectively for any of the taboo probability and distribution functions as defined in § II.11 (rather than § II.12); we put

$$\hat{p}(x) = \int_0^\infty e^{-xt}\, p(t)\, dt; \qquad \hat{F}(x) = \int_0^\infty e^{-xt}\, dF(t);$$

$$P(t) = \int_0^t p(s)\, ds; \qquad \hat{P}(x) = \int_0^\infty e^{-xt}\, P(t)\, dt = x^{-1}\hat{p}(x).$$

These Laplace transforms are defined for $x > 0$ and all subsequent formulas involving them are valid there. The conventions regarding ∞ as a "number" given in § I.9 will be used.

A special case of (11.19) is as follows:

$$(1) \qquad p_{ii}(t) = {}_jp_{ii}(t) + \int_0^t p_{ji}(t-s)\, dF_{ij}(s).$$

Taking Laplace transforms we have

$$(2) \qquad \hat{p}_{ii}(x) = {}_j\hat{p}_{ii}(x) + \hat{p}_{ji}(x)\, \hat{F}_{ij}(x).$$

Another special case of (11.19) is as follows:

$$(3) \qquad p_{ij}(t) = \delta_{ij}\, e^{-q_i t} + \int_0^t p_{jj}(t-s)\, dF_{ij}(s).$$

Letting $i \neq j$ and taking Laplace transforms we have

$$(4) \qquad \hat{p}_{ij}(x) = \hat{p}_{jj}(x)\, \hat{F}_{ij}(x).$$

By a simple Abelian theorem, we have

$$\lim_{x\to 0} \hat{p}_{jj}(x) = P_{jj}(\infty); \qquad \lim_{x\to 0} \hat{F}_{ij}(x) = F_{ij}(\infty).$$

It follows directly from (3) that if $i \neq j$ then

$$(5) \qquad \lim_{t\to\infty} \frac{P_{ij}(t)}{P_{jj}(t)} = F_{ij}(\infty).$$

Next, interchanging i and j in (4) and substituting into (2), we obtain

$$(6) \qquad \hat{p}_{ii}(x)\,[1 - \hat{F}_{ij}(x)\, \hat{F}_{ji}(x)] = {}_j\hat{p}_{ii}(x).$$

It follows from (6) that

$$(7) \qquad P_{ii}(\infty) = \frac{{}_jP_{ii}(\infty)}{1 - F_{ij}(\infty)\, F_{ji}(\infty)}.$$

If $F_{ij}(\infty) = 0$ then (1) reduces to ${}_jp_{ii}(t) \equiv p_{ii}(t)$ so that (7) is trivial. From (5) we see that if $i \neq j$, then $i \rightsquigarrow j$ if and only if $F_{ij}(\infty) > 0$; moreover if $i \rightsquigarrow j$ then ${}_jP_{ii}(\infty) < \infty$ by Theorem 11.5. On the other hand

${}_jP_{ii}(\infty) > 0$ by the remarks preceding Theorem 11.2. Hence we conclude from (7) and Theorem 10.4 as follows.

Theorem 1. The state i is recurrent if and only if $F_{ij}(\infty)F_{ji}(\infty) = 1$ for every j such that $i \rightsquigarrow j$. Furthermore if i is nonrecurrent then $F_{ij}(\infty)F_{ji}(\infty) < 1$ for every j and (7) holds for every $j \neq i$.

Next, let us introduce the *mean first entrance time from i to j*:

$$m_{ij} = \int_0^\infty t\, dF_{ij}(t);$$

if $i = j$ this is the *mean first return time of i*, which is zero if i is instantaneous. A Tauberian theorem by Hardy and Littlewood (see e.g. Hardy [1; Theorem 98]) asserts that

$$\lim_{t\to\infty} \frac{1}{t} P_{ii}(t) = \lim_{x\to 0} x\hat{p}_{ii}(x).$$

It follows from (6) and l'Hospital's rule that

$$\lim_{x\to 0} x\hat{p}_{ii}(x) = \frac{{}_jP_{ii}(\infty)}{m_{ij} + m_{ji}}.$$

Theorem 10.1 together with the last two formulas yield the following theorem.

Theorem 2. If i is recurrent, and $i \rightsquigarrow j \neq i$, then

$$\lim_{t\to\infty} p_{ii}(t) = \frac{{}_jP_{ii}(\infty)}{m_{ij} + m_{ji}}. \tag{8}$$

Corollary. If i is recurrent then i is positive or null according as $m_{ij} + m_{ji}$ is finite or infinite for every $j \leftrightsquigarrow i$, $j \neq i$.

Theorem 2 and its corollary are valid for an arbitrary i. For a stable i there are simpler results which do not involve another state, as we now proceed to derive. Putting $i = j$ in (1) and (3) we have

$$p_{ii}(t) = e^{-q_i t} + \int_0^t p_{ii}(t-s)\, dF_{ii}(s).$$

Taking Laplace transforms we obtain

$$(q_i + x)\, \hat{p}_{ii}(x)\, [1 - \hat{F}_{ii}(x)] = 1,$$

in place of (6). The rest follows as before and is stated in the following theorem.

Theorem 3. Let i be stable. Then i is recurrent if and only if $F_{ii}(\infty) = 1$; if i is nonrecurrent we have

$$P_{ii}(\infty) = \frac{1}{q_i[1 - F_{ii}(\infty)]}.$$

In the recurrent case we have

$$\lim_{t\to\infty} p_{ii}(t) = \frac{1}{q_i m_{ii}}. \tag{9}$$

These are therefore the exact analogues of Theorems I.5.4 and I.6.1. Observe that the limit in (9) can also be written in terms of the random variables introduced in § 11 as

$$\frac{\boldsymbol{E}(\varrho_i \mid \Delta_i)}{\boldsymbol{E}(\alpha_{ii} \mid \Delta_i)},$$

which is the average time spent in the state i between two successive entrances into i. We may now proceed to establish results analogous to those of § I.9; but we will concern ourselves only with the crucial Theorem I.9.5.

In the following lemma the Laplace transform $\hat{A}$ of A is defined by $\hat{A}(x) = \int_0^\infty e^{-xt}\, dA(t)$; similarly for B.

Lemma. Let A be a nondecreasing function on $[0, \infty)$ such that $0 = A(0) \leqq A(0+) < 1$, $A(\infty) \leqq 1$. There exists a finite nondecreasing function B on $[0, \infty)$ such that $B(t+1) - B(t)$ does not exceed a constant independent of t, and such that the Laplace transforms are related as follows:

$$\frac{1}{1-\hat{A}(x)} = \hat{B}(x), \qquad x > 0.$$

For a simple probabilistic proof of the lemma see Blackwell [1; Lemma 3]. It can be shown, though we do not need it, that if A is continuous so is B.

Let $i \neq j$, $i \rightsquigarrow j$ and $A = F_{ij} * F_{ji}$ where $*$ denotes the usual convolution of distribution functions. Then $A(0+) = F_{ij}(0) F_{ji}(0) = 0$ by Theorem 6.2. Since $i \rightsquigarrow j$ we have $0 < A(\infty) \leqq 1$. Hence the conditions of the Lemma are satisfied. Using the function B given by the lemma we may therefore write (6) as

$$\hat{p}_{ii}(x) = {}_j\hat{p}_{ii}(x)\, \hat{B}(x).$$

It follows from the uniqueness theorem for Laplace transforms (see e.g. Widder [1; p. 63]) that

$$p_{ii}(t) = \int_0^t {}_j p_{ii}(t-s)\, dB(s),$$

and consequently upon integration

$$P_{ii}(t) = \int_0^t {}_j P_{ii}(t-s)\, dB(s). \tag{10}$$

Interchanging i and j and observing that A is symmetric in i and j so that B is unchanged, we have

$$P_{jj}(t)=\int_0^t {}_iP_{jj}(t-s)\,dB(s). \tag{11}$$

By the remarks after (7) we have if $i \neq j$, $0<{}_jP_{ii}(\infty)<\infty$, $0<{}_iP_{jj}(\infty)<\infty$. It now follows by elementary analysis from (10), (11) and the bounded increment property of B that, regardless of the finiteness of $B(\infty)$,

$$\lim_{t\to\infty}\frac{P_{ii}(t)}{P_{jj}(t)}=\frac{{}_jP_{ii}(\infty)}{{}_iP_{jj}(\infty)}. \tag{12}$$

Theorem 4. If $i \leftrightsquigarrow j$ then (12) holds, the limit being positive and finite.

For another proof of (12) see after Theorem 8.

To obtain further expressions for the limit in (12) in the case where at least one of i and j is stable, we proceed as follows. If i is stable and $i \neq j$, we have as special cases of (11.19) and (11.20):

$${}_jp_{ii}(t)=e^{-q_i t}+\int_0^t {}_jp_{ii}(t-s)\,d\,{}_jF_{ii}(s); \tag{13}$$

$${}_ip_{ij}(t)=\int_0^t {}_ip_{jj}(t-s)\,d\,{}_iF_{ij}(s); \tag{14}$$

$$F_{ij}(t)={}_iF_{ij}(t)+\int_0^t F_{ij}(t-s)\,d\,{}_jF_{ii}(s). \tag{15}$$

Integrating (13) and (14) over $[0,\infty)$ we obtain

$${}_jP_{ii}(\infty)=\frac{1}{q_i[1-{}_jF_{ii}(\infty)]}; \tag{16}$$

$${}_iP_{ij}(\infty)={}_iP_{jj}(\infty)\,{}_iF_{ij}(\infty). \tag{17}$$

Letting $t\to\infty$ in (15) we obtain

$$F_{ij}(\infty)=\frac{{}_iF_{ij}(\infty)}{1-{}_jF_{ii}(\infty)}. \tag{18}$$

If now $i \leftrightsquigarrow j$ so that all the probabilities involved here are finite and positive, it follows that

$$\frac{{}_jP_{ii}(\infty)}{{}_iP_{jj}(\infty)}=\frac{F_{ij}(\infty)}{q_i\,{}_iP_{ij}(\infty)}; \tag{19}$$

if furthermore j is also stable then the above is also equal to

$$\frac{q_j F_{ij}(\infty)\,{}_jF_{ji}(\infty)}{q_i F_{ji}(\infty)\,{}_iF_{ij}(\infty)}. \tag{20}$$

Theorem 5. If i is stable, $i \neq j$ and $i \rightsquigarrow j$, the limit in (12) is also given by (19); if in addition j is stable it is furthermore given by (20).

We add another formula involving three states, which will be used in § 14. Let $i \neq j$, $i \rightsquigarrow j \rightsquigarrow k$. A special case of (11.19) is

$$p_{ik}(t) = {}_j p_{ik}(t) + \int_0^t p_{jk}(t-s)\, dF_{ij}(s);$$

or passing to Laplace transforms and dropping the x,

$$\hat{p}_{ik} = {}_j\hat{p}_{ik} + \hat{p}_{jk}\hat{F}_{ij}. \tag{21}$$

Interchanging i and j and substituting the resulting formula into (21), we obtain

$$\hat{p}_{ik}(1 - \hat{F}_{ij}\hat{F}_{ji}) = \hat{F}_{ij}\,{}_i\hat{p}_{jk} + {}_j\hat{p}_{ik}.$$

This is a generalization of (6). Furthermore it follows from (4) and (6) that if $i \neq k$,

$$\hat{p}_{ik} = \hat{F}_{ik}\hat{p}_{kk} = \hat{F}_{ik}\frac{{}_i\hat{p}_{kk}}{{}_k\hat{p}_{ii}}\,\frac{{}_j\hat{p}_{ii}}{1 - \hat{F}_{ij}\hat{F}_{ji}}.$$

Comparing the last two formulas we obtain

$$\hat{F}_{ik}\frac{{}_i\hat{p}_{kk}}{{}_k\hat{p}_{ii}}\,{}_j\hat{p}_{ii} = \hat{F}_{ij}\,{}_i\hat{p}_{jk} + {}_j\hat{p}_{ik},$$

from which the following theorem is derived by letting $x \to 0$.

Theorem 6. If $i \neq j$ and i, j and k belong to one recurrent class then

$$\frac{{}_iP_{kk}(\infty)\,{}_jP_{ii}(\infty)}{{}_kP_{ii}(\infty)} = {}_iP_{jk}(\infty) + {}_jP_{ik}(\infty). \tag{22}$$

Observe that the results in this section were derived from the first entrance decomposition of § 11 without using the last exit decomposition of § 12. As a matter of fact, they were obtained before the results of § 12, which are more difficult, were proved. The latter may be used to yield quicker proofs as well as new formulas. For instance, we deduce from Theorem 12.13 or just (12.24) that

$$\lim_{t\to\infty}\frac{P_{ij}(t)}{P_{ii}(t)} = G_{ij}(\infty)$$

as the dual to (5); and therefore

$$\lim_{t\to\infty}\frac{P_{ii}(t)}{P_{jj}(t)} = \frac{F_{ij}(\infty)}{G_{ij}(\infty)}$$

as a complement to Theorem 4, the finiteness of $G_{ij}(\infty)$ in this case being a consequence.

In what sense and how well do the discrete skeletons $\mathfrak{C}_s$ approximate the M.C. $\mathfrak{C}$ as $s\downarrow 0$? We have already seen on several occasions, notably in Theorems 10.2 and 10.4, that results about a d.p. M.C. can be used to obtain the corresponding ones about a c.p. M.C. by suitable approximations. Further illustrations of this method will now be given.

Consider the discrete skeletons $\mathfrak{C}_s$: $\{x_{ns},\ n\geqq 0\}$ of $\mathfrak{C}$: $\{x_t,\ t\geqq 0\}$. Quantities defined for a skeleton will be denoted by the symbols used for a d.p. M.C. in Part I (see in particular § I.9) followed by a parenthesis indicating the scale of the skeleton; e.g.

$${}_Hp_{ij}^{(n)}(s)=\boldsymbol{P}\{x_{\nu s}(\omega)\notin H,\ 1\leqq\nu<n;\ x_{ns}(\omega)=j\,|\,x_0(\omega)=i\}$$

$${}_Hp_{ij}^{*}(s)=\sum_{n=1}^{\infty}{}_Hp_{ij}^{(n)}(s)\,;\qquad {}_Hf_{ij}^{(n)}(s)={}_{j,H}p_{ij}^{(n)}(s)\,;$$

$$m_{ij}(s)=\sum_{n=1}^{\infty}n f_{ij}^{(n)}(s)\,.$$

It is clear that ${}_Hp_{ij}(t)$ is a good analogue of ${}_Hp_{ij}^{(n)}$ only if i is instantaneous while we need other definitions if i is stable. Note also that although $p_{ij}^{(n)}(s)=p_{ij}(ns)$, ${}_Hp_{ij}^{(n)}(s)$ is not equal to ${}_Hp_{ij}(ns)$ in general. We shall suppose $i\notin H$ below so that the definitions of the latter given in § 11 and § 12 agree.

Let us take the x_+ version in what follows and define on $\Delta_i=\{\omega\colon x_0(\omega)=i\}$ for each $s>0$,

$$\varrho_i^{(s)}(\omega)=\max\{ns\colon n\geqq 0;\ x(\nu s,\omega)=i,\ 0\leqq\nu\leqq n\}.$$

Furthermore we define for each $s>0$,

$$\alpha_{ij}^{(s)}(\omega)=\min\{ns\colon n\geqq 1;\ ns>\varrho_i^{(s)}(\omega);\ x(ns,\omega)=j\},$$

$$\beta_{ij}^{(s)}(\omega)=\min\{ns\colon n\geqq 1;\ ns\geqq\alpha_{ij}(\omega);\ x(ns,\omega)=j\},$$

if $\omega\in\Delta_{ij}=\Delta_i\cap\{\omega\colon\alpha_{ij}(\omega)<\infty\}$; and $\alpha_{ij}^{(s)}(\omega)=\beta_{ij}^{(s)}(\omega)=\infty$ if $\omega\in\Delta_i-\Delta_{ij}$.

For some applications it would actually be more expedient to define $\varrho_i^{(s)}(\omega)\equiv 0$ if i is instantaneous, but the definitions given above lead to more interesting results in this case. The random variables ϱ_i and α_{ij} are defined in (11.1) and (11.12).

Theorem 7. We have

$$\boldsymbol{P}\{\lim_{s\downarrow 0}\varrho_i^{(s)}(\omega)=\varrho_i(\omega)\,|\,\Delta_i\}=1\,. \tag{23}$$

If j is stable we have

$$\boldsymbol{P}\{\lim_{s\downarrow 0}\alpha_{ij}^{(s)}(\omega)=\lim_{s\downarrow 0}\beta_{ij}^{(s)}(\omega)=\alpha_{ij}(\omega)\,|\,\Delta_{ij}\}=1\,. \tag{24}$$

If j is instantaneous, (24) holds if the limits are taken along any given sequence $\{s_\nu\} \to 0$; but for a.a. ω in Δ_{ij} the set of s for which $\alpha_{ij}^{(s)}(\omega)$ or $\beta_{ij}^{(s)}(\omega)$ is finite is of the first category.

Proof. If i is absorbing then $\varrho_i^{(s)}(\omega) = \varrho_i(\omega) = \infty$ for a.a. ω in Δ_i. If i is stable but not absorbing, (23) follows from Theorem 5.7. In fact, for a.a. ω in Δ_i, there exists an $s_0(\omega)$ such that $x(t, \omega) \neq i$ for $\varrho_i(\omega) < t < \varrho_i(\omega) + s_0(\omega)$ so that $\varrho_i(\omega) - s \leq \varrho_i^{(s)}(\omega) \leq \varrho_i(\omega)$ if $s < s_0(\omega)$. If i is instantaneous then $\varrho_i(\omega) = 0$ for a.a. ω in Δ_i by (11.2); hence it is sufficient to prove that $\lim\limits_{s \downarrow 0} \varrho_i^{(s)}(\omega) = 0$ for a.a. ω in Δ_i. Assume the contrary, then there is a set $\Lambda \subset \Delta_i$ with $\boldsymbol{P}(\Lambda) > 0$ and a constant $c > 0$ such that if $\omega \in \Lambda$ then $\overline{\lim\limits_{s \downarrow 0}}\, \varrho_i^{(s)}(\omega) > c$. Hence there exists a sequence $\{s_\nu(\omega)\}$ decreasing to zero such that $x(n s_\nu(\omega), \omega) = i$ for all $n s_\nu(\omega)$ in the interval $(0, c)$. It follows that $S_i(\omega)$ is dense there. This is impossible by Theorem 6.6, proving (23).

Turning to (24) it is easy to see by Theorem 5.6 that if j is stable but not absorbing, then for a.a. ω in Δ_{ij} there exists a $\delta(\omega) > 0$ such that if $s < \delta(\omega)$ then firstly $\varrho_i^{(s)}(\omega) \leq \alpha_{ij}(\omega)$ so that $\alpha_{ij}^{(s)}(\omega) = \beta_{ij}^{(s)}(\omega)$ and secondly $\alpha_{ij}^{(s)}(\omega) < \alpha_{ij}(\omega) + s$. Hence (24) is true.

Next let j be instantaneous and assume that there exists a sequence $\{s_\nu, \nu \geq 1\}$ decreasing to zero, a set $\Lambda \subset \Delta_{ij}$ with $\boldsymbol{P}(\Lambda) > 0$, and a constant $c > 0$ such that if $\omega \in \Lambda$ then $\beta_{ij}^{(s_\nu)}(\omega) > \alpha_{ij}(\omega) + c$ for all ν. Since the set of numbers $\{n s_\nu, n \geq 1, \nu \geq 1\}$ is everywhere dense in $\boldsymbol{T}$ it is a separability set and we have $x(n s_\nu, \omega) \neq j$ for all $n s_\nu$ in the interval $(\alpha_{ij}(\omega), \alpha_{ij}(\omega) + c)$. This is impossible by the definition of α_{ij} and Theorem 4.1. Hence $\lim\limits_{\nu \to \infty} \beta_{ij}^{(s_\nu)}(\omega) = \alpha_{ij}(\omega)$ for a.a. ω. The proof for $\alpha_{ij}^{(s_\nu)}$ is similar if we observe that $\alpha_{ij}^{(s)} \geq \alpha_{ij}$ whether $i = j$ or not.

It remains to prove the last assertion of the theorem. For an instantaneous j and a.a. ω the set $S_j(\omega)$ is nowhere dense in $\boldsymbol{T}$. Denoting the set $\{t: nt \in S_j(\omega)\}$ by $n^{-1} S_j(\omega)$ we see that it is likewise nowhere dense and hence the union $U(\omega) = \bigcup\limits_{n=1}^{\infty} [n^{-1} S_j(\omega)]$ is a set of the first category (see e.g. KELLEY [1; p. 201]). If $s \in \boldsymbol{T} - U(\omega)$ we have $ns \notin S_j(\omega)$ for all $n \geq 1$ and hence by definition $\alpha_{ij}^{(s)}(\omega) = \beta_{ij}^{(s)}(\omega) = \infty$. In particular $\overline{\lim\limits_{s \downarrow 0}}\, \alpha_{ij}^{(s)}(\omega) = \infty$ since $\boldsymbol{T} - U(\omega)$ has 0 as a limit point. ▯

Corollary. We have

$$(25) \qquad \lim_{s \downarrow 0} \boldsymbol{P}\{\alpha_{ij}^{(s)}(\omega) \leq \alpha_{ik}^{(s)}(\omega) \mid \Delta_{ij}\} = \boldsymbol{P}\{\alpha_{ij}(\omega) \leq \alpha_{ik}(\omega) \mid \Delta_{ij}\}.$$

Proof. If $j \neq k$ we have $\boldsymbol{P}\{\alpha_{ij}(\omega) = \alpha_{ik}(\omega)\} = 0$ by Theorem 7.5. The corollary then follows from the theorem if $\omega \in \Delta_{ij} \cap \Delta_{ik}$ since the relation (24) for every given sequence $\{s_\nu\} \to 0$ implies the corresponding convergence in probability; otherwise it is trivial. ▯

Theorem 8. Suppose that H is a finite taboo set, $i \notin H$ and $i \neq j$; then we have for $0 \leqq t \leqq \infty$:

$$\lim_{s\downarrow 0} \sum_{n=1}^{[ts^{-1}]} {}_Hf_{ij}^{(n)}(s) = {}_HF_{ij}(t). \tag{26}$$

Furthermore, if $j \notin H$ and $j \rightsquigarrow H$, then we have for $0 \leqq t \leqq \infty$:

$$\lim_{s\downarrow 0} \sum_{n=0}^{[ts^{-1}]} {}_Hp_{ij}^{(n)}(s) = \int_0^t {}_Hp_{ij}(u)\,du. \tag{27}$$

Thus in particular for $t = \infty$:

$$\lim_{s\downarrow 0} {}_Hf_{ij}^{*}(s) = {}_HF_{ij}(\infty); \tag{28}$$

$$\lim_{s\downarrow 0} s\, {}_Hp_{ij}^{*}(s) = {}_HP_{ij}(\infty). \tag{29}$$

Proof. The sum in the left member of (26) is equal to

$$\boldsymbol{P}\{\alpha_{ij}^{(s)}(\omega) \leqq t;\ \alpha_{ij}^{(s)}(\omega) \leqq \min_{k\in H} \alpha_{ij}^{(s)}(\omega) \mid \varDelta_i\},$$

while the right member of (26) is equal to

$$\boldsymbol{P}\{\alpha_{ij}(\omega) \leqq t;\ \alpha_{ij}(\omega) \leqq \min_{k\in H} \alpha_{ik}(\omega) \mid \varDelta_i\}.$$

Hence (26) follows from (25). Next we have

$$ {}_Hp_{ij}^{(n)}(s) = \boldsymbol{P}\left\{\min_{k\in H} \alpha_{ik}^{(s)}(\omega) \geqq ns;\ x(ns, \omega) = j \mid \varDelta_i\right\}. \tag{30}$$

It follows from Theorem 7, the stochastic continuity of $\{x_t, t \in \boldsymbol{I}\}$ and (11.14) that

$$\lim_{s\downarrow 0} {}_Hp_{ij}^{([us^{-1}])}(s) = \boldsymbol{P}\left\{\min_{k\in H} \alpha_{ik}(\omega) \geqq u;\ x(u, \omega) = j \mid \varDelta_i\right\} = {}_Hp_{ij}(u). \tag{31}$$

Since

$$s \sum_{n=0}^{[ts^{-1}]} {}_Hp_{ij}^{(n)}(s) = \int_0^{[ts^{-1}]s+s} {}_Hp_{ij}^{([us^{-1}])}(s)\,du,$$

(27) follows from (31) by bounded convergence if $t < \infty$. To prove (27) when $t = \infty$ we shall show that the convergence of the left member of (27) to $s\, {}_Hp_{ij}^{*}(s)$ as $t \to \infty$ is uniform in $s \in (0, s_0)$ for some $s_0 > 0$. That is, given any $\varepsilon > 0$, there exists a $t = t(\varepsilon)$ such that

$$s \sum_{n > ts^{-1}} {}_Hp_{ij}^{(n)}(s) < \varepsilon$$

for $0 < s < s_0$. For this purpose we re-examine the proof of Theorem I.9.3. Let $k \in H$ and $p_{jk}(t_0) > 0$. Since p_{jk} is continuous, there is a $\delta > 0$ and a $t_1 > t_0$ such that $p_{jk}(t) \geqq \delta$ if $t_0 \leqq t \leqq t_1$. For each sufficiently small s there exists an integer m_s such that $t_0 \leqq m_s s \leqq t_1$. We have

$$s \sum_{n > ts^{-1}} {}_H p_{ij}^{(n)}(s) \leqq \frac{s}{p_{jk}(m_s s)} \sum_{n > ts^{-1}} \boldsymbol{P}\{x(\nu s, \omega) \neq k,$$
$$1 \leqq \nu \leqq n;\ x\big((n+m_s)s, \omega\big) = k \mid \varDelta_i\}.$$

The last sum may be written, by putting $n = \lambda m_s + r$, as

$$\sum_{r=0}^{m_s-1} \sum_{\lambda > (ts^{-1}-r)m_s^{-1}} \boldsymbol{P}\{x(\nu s, \omega) \neq k,\ 1 \leqq \nu \leqq \lambda m_s + r;$$
$$x\big((\lambda m_s + r + m_s)s, \omega\big) = k \mid \varDelta_i\}$$
$$\leqq m_s \boldsymbol{P}\{x(\nu s, \omega) \neq k,\ 1 \leqq \nu \leqq [ts^{-1}] + 1 \mid \varDelta_i\}$$
$$\leqq m_s \boldsymbol{P}\{\alpha_{ik}^{(s)}(\omega) > t \mid \varDelta_i\}.$$

It follows from Theorem 8 that

$$\overline{\lim_{s\downarrow 0}}\, s \sum_{n > ts^{-1}} {}_H p_{ij}^{(n)}(s) \leqq \overline{\lim_{s\downarrow 0}} \frac{m_s s}{p_{jk}(m_s s)} \int_t^\infty dF_{ik}(u) \leqq \frac{t_1}{\delta} \int_t^\infty dF_{ik}(u).$$

The right member being independent of s, this proves the asserted uniform convergence and therefore (29). ▯

As a first application of the results above, we give a one-line proof of (12). By (10.9), Corollary 1 to Theorem I.9.4 and (29) above, we have

$$\lim_{t\to\infty} \frac{P_{ii}(t)}{P_{jj}(t)} = \lim_{s\downarrow 0} \lim_{N\to\infty} \frac{\sum_{n=0}^{N} p_{ii}^{(n)}(s)}{\sum_{n=0}^{N} p_{jj}^{(n)}(s)} = \lim_{s\downarrow 0} \frac{s\, {}_j p_{ii}^*(s)}{s\, {}_i p_{jj}^*(s)} = \frac{{}_j P_{ii}(\infty)}{{}_i P_{jj}(\infty)}.$$

Our next application deals with mean first entrance and return times.

Theorem 9. If i and j belong to a positive class and $i \neq j$, then

$$\lim_{s\downarrow 0} s\, m_{ij}(s) = m_{ij}. \tag{32}$$

If i is stable we have

$$m_{ii}(s) = q_i\, m_{ii}. \tag{33}$$

Proof. (32) follows immediately from (9) and Theorem I.6.1. Next, we have from Corollary 2 to Theorem I.11.3,

$$m_{ij}(s) + m_{ji}(s) = m_{ii}(s)\,[1 + {}_j p_{ii}^*(s)].$$

Hence it follows from Theorem I.6.1. that

$$\lim_{n\to\infty} p_{ii}^{(n)}(s) = \frac{1}{m_{ii}(s)} = \frac{1 + {}_j p_{ii}^*(s)}{m_{ij}(s) + m_{ji}(s)}.$$

This being the same limit as $\lim_{t\to\infty} p_{ii}(t)$ we obtain by comparison with (8):

$$\frac{1+{}_jp^*_{ii}(s)}{m_{ij}(s)+m_{ji}(s)}=\frac{{}_jP_{ii}(\infty)}{m_{ij}+m_{ji}}.$$

It now follows from (29) applied to the numerators above that

$$\lim_{s\downarrow 0} s\,[m_{ij}(s)+m_{ji}(s)]=m_{ij}+m_{ji}. \tag{34}$$

On the other hand we have from (26) and (28),

$$\lim_{s\downarrow 0}\sum_{n\geqq[ts^{-1}]+1} f_{ij}^{(n)}(s)=1-F_{ij}(t).$$

Hence we have by Fatou's lemma,

$$\varliminf_{s\downarrow 0}\int_0^\infty \sum_{n\geqq[ts^{-1}]+1} f_{ij}^{(n)}(s)\,dt\geqq\int_0^\infty [1-F_{ij}(t)]\,dt=m_{ij}.$$

The integral in the first member above is equal to

$$\sum_{m=1}^{\infty} s\sum_{n\geqq m} f_{ij}^{(n)}(s)=s\,m_{ij}(s).$$

Thus we have

$$\varliminf_{s\downarrow 0} s\,m_{ij}(s)\geqq m_{ij}. \tag{35}$$

(32) follows from (34) and (35). □

We turn to the analogue of Theorem I.9.7. Suppose that all states form one recurrent class. If there is a stable state h then the system $\{{}_hP_{hk}(\infty),\ k\in\boldsymbol{I}\}$ has properties similar to $\{{}_hp^*_{hk},\ k\in\boldsymbol{I}\}$ in a d.p. M.C. If there is no stable state then there are infinitely many states; let $i\neq j$. We shall see that the system $\{e_k^{(i,j)},\ k\in\boldsymbol{I}\}$ where

$$e_k^{(i,j)}={}_jP_{ik}(\infty)+{}_iP_{jk}(\infty) \tag{36}$$

takes the place of the preceding system.

Theorem 10. Let $\boldsymbol{I}$ be a recurrent class. Then for each pair (i,j), $i\neq j$, the system $\{e_k^{(i,j)},\ k\in\boldsymbol{I}\}$ defined by (36) is a solution of the system of equations

$$u_l=\sum_k u_k\,p_{kl}(s),\qquad l\in\boldsymbol{I} \tag{37}$$

in the unknowns $\{u_k,\ k\in\boldsymbol{I}\}$, for every $s>0$. Conversely, any solution of (37) for a sequence of values of s tending to zero is proportional to the system $\{e_k^{(i,j)},\ k\in\boldsymbol{I}\}$.

Remark. If the system (37) holds for a certain value of s it holds for all larger values.

Proof. Let us verify for the discrete skeleton $\mathfrak{C}_s$ that $\{e_k^{(i,j)}(s),\ k\in\boldsymbol{I}\}$, where

$$e_k^{(i,j)}(s) = {}_jp_{ik}^*(s) + {}_ip_{jk}^*(s),$$

is a solution of the system (37). We have

$$\begin{aligned}\sum_k \{{}_jp_{ik}^*(s) + {}_ip_{jk}^*(s)\}p_{kl}^{(1)}(s) &= \sum_{n=1}^{\infty}\Big\{\sum_k {}_jp_{ik}^{(n)}(s)\,p_{kl}^{(1)}(s) + \sum_k {}_ip_{jk}^{(n)}(s)\,p_{kl}^{(1)}(s)\Big\}\\ &= \sum_{n=1}^{\infty}\{{}_jp_{il}^{(n+1)}(s) + {}_jp_{ij}^{(n)}(s)\,p_{jl}^{(1)}(s) + {}_ip_{jl}^{(n+1)}(s) + {}_ip_{ji}^{(n)}(s)\,p_{il}^{(1)}(s)\}\\ &= {}_jp_{il}^*(s) - {}_jp_{il}^{(1)}(s) + {}_jp_{ij}^*(s)\,p_{jl}^{(1)}(s) + {}_ip_{jl}^*(s) - {}_ip_{jl}^{(1)}(s) + {}_ip_{ji}^*(s)\,p_{il}^{(1)}(s).\end{aligned}$$

Since ${}_jp_{ij}^*(s) = {}_ip_{ji}^*(s) = 1$ in a recurrent class and ${}_jp_{il}^{(1)}(s) = p_{il}^{(1)}(s)$, ${}_ip_{jl}^{(1)}(s) = p_{jl}^{(1)}(s)$, this reduces to ${}_jp_{il}^*(s) + {}_ip_{jl}^*(s)$, verifying (37). Thus,

$$e_l^{(i,j)}(s) = \sum_k e_k^{(i,j)}(s)\,p_{k\,l}(s).$$

It follows by iteration (in $\mathfrak{C}_s$) that for any positive integer m,

$$e_l^{(i,j)}(s) = \sum_k e_k^{(i,j)}(s)\,p_{k\,l}(ms).$$

This being true for every s, we have any given $t>0$,

$$e_l^{(i,j)}(t/m) = \sum_k e_k^{(i,j)}(t/m)\,p_{k\,l}(t).$$

Applying Theorem I.9.7 to the skeleton $\mathfrak{C}_t$, we have for a fixed h and constant c_m:

$$e_l^{(i,j)}\left(\frac{t}{m}\right) = c_m\,{}_hp_{hl}^*(t).$$

By (29),

$$\lim_{m\to\infty}\frac{t}{m}\,e_l^{(i,j)}\left(\frac{t}{m}\right) = e_l^{(i,j)}.$$

Hence $c = \lim\limits_{m\to\infty} c_m/m$ exists and

$$e_l^{(i,j)} = ct\,{}_hp_{hl}^*(t).$$

It follows from Theorem I.9.7 that

$$e_l^{(i,j)} = \sum_k e_k^{(i,j)}\,p_{k\,l}(t), \tag{38}$$

proving the first part of the theorem.

Conversely, let (37) hold for $s=s_n$ where $s_n\downarrow 0$. Since $\{e_k^{(i,j)}(s_n),\ k\in\boldsymbol{I}\}$ is also a solution it follows from Theorem I.9.7 that the two solutions must be proportional, namely:

$$u_k = c(s_n)\,s_n\,e_k^{(i,j)}(s_n)$$

where $c(s_n)$ does not depend on k. Letting $s_n\downarrow 0$ we see from (29) that $\lim\limits_{n\to\infty} c(s_n) = c$ exists and $u_k = ce_k^{(i,j)}$ as asserted. ▯

As a consequence of Theorem 5, for given (i, j) and (i', j') where $i \neq j$, $i' \neq j'$, the ratio $e_k^{(i,j)}/e_k^{(i',j')}$ does not depend on k; for a direct proof see the next section. It is also clear from the proof of Theorem 10 that if "$=$" is replaced by "$\geqq$" in (38), then the result holds in any $\boldsymbol{I}$.

Notes. This section is merged from § 12 and § 13 of the first edition to make room for the new § 12. Theorem 4 was first proved in CHUNG [1] under a severely restrictive condition which makes almost all sample functions step functions (see Theorem 19.2). The analytic part of the proof was the same; it is the relevant part of Theorem 11.8 that was lacking. No purely analytic proof of the existence of the limit in (12) seems known. The usual ratio Tauberian theorem does not apply; cf. § I.10 and see also CHUNG [7]. The last-cited paper contains most of the results up to but not including Theorem 6. As remarked in the text, the apparatus of § 12 has not been employed here.

The results on discrete approximations, beginning with Theorem 7, are new and serve only to indicate some of the possibilities. The extension of some of the discrete parameter results such as Theorem 11.1 seems straight-forward; on the other hand, matters concerning the approximation of corresponding discrete and continuous parameter quantities can be quite delicate and much remains to be done. When one reflects that separability is a basic assumption in the continuous parameter case, and that the denumerable set of separability may be obtained by successive refinement (see e.g., the proof of Theorem 5.5), the intricate limiting process involved in the passage from the discrete to the continuous should offer many a challenging question indeed.

§ 14. Functionals

In this section we indicate how the developments of Part I, §§ 14 to 16 can be extended to the continuous parameter case. The main idea of recurrence to a fixed state and the consequent sectioning of the time axis has an obvious analogue if there is a stable state. In the general case this will be replaced by the alternation of two distinct instantaneous states.

Let $\boldsymbol{I}$ be a recurrent class; and put

$$\frac{{}_jP_{ii}(\infty)}{{}_iP_{jj}(\infty)} = d_{ij}.$$

By Theorem 13.4 we have

$$d_{ij}\, d_{jk}\, d_{ki} = \lim_{t\to\infty} \frac{P_{ii}(t)\, P_{jj}(t)\, P_{kk}(t)}{P_{jj}(t)\, P_{kk}(t)\, P_{ii}(t)} = 1$$

and consequently

$$d_{ij}\, d_{jk} = d_{ik}.$$

In particular if h is a fixed state distinct from i and j, then $d_{ij}=d_{ih}\,d_{hj}$. Thus d_{ij} as a function of i and j is the product of two numbers, one of which depends only on i and the other only on j. Formula (12.22) may be written as

$$e_k^{(i,j)}={}_jP_{ii}(\infty)\,d_{ki}=\left({}_jP_{ii}(\infty)\,d_{hi}\right)d_{kh}.$$

Thus $e_k^{(i,j)}$ as a function of (i,j,k) is the product of a symmetric function of (i,j) alone and a function of k alone, confirming the remarks at the end of § 13.

We have already seen in Theorem 13.10 that the system $\{e_k^{(i,j)},\ k\in\boldsymbol{I}\}$ plays a role similar to that of $\{e_{hk},\ k\in\boldsymbol{I}\}$ in § I.9. The exact analogue of $\{e_k^{(i,j)}\}$ is $\{{}_jp_{ik}^*+{}_ip_{jk}^*\}$ which is seen by the proof of Theorem 13.10 and Theorem I.9.7 to be proportional to $\{e_{hk}\}$ for fixed i, j and h. Indeed we could have used the former instead of the latter in the discrete parameter case but this was unnecessary since the simpler $\{e_{hk}\}$ sufficed. In the continuous parameter case, if there is a stable state h, then $\{q_h\,{}_hP_{hk}(\infty),\ k\in\boldsymbol{I}\}$ is the exact analogue of $\{e_{hk},\ k\in\boldsymbol{I}\}$ and can be used for the same purposes. In general however the system $\{e_k^{(i,j)},\ k\in\boldsymbol{I}\}$ is to be used, as we shall see presently.

Let f be a single-valued function from $\bar{\boldsymbol{I}}$ to $(-\infty,\infty)$. Let

$$y_t(\omega)=y(t,\omega)=f[x(t,\omega)].$$

Then $\{y_t,\ t\in\boldsymbol{T}\}$ is called a *functional* of the M.C. $\{x_t,\ t\in\boldsymbol{T}\}$. Clearly $\{y_t\}$ is a stochastic process; if $\{x_t\}$ is measurable (or Borel measurable) so is $\{y_t\}$. However $\{y_t\}$ need not be a M.C.; it is if f is one-to-one. In what follows the M.C. $\{x_t\}$ is assumed to be at least separable and measurable, the separability of $\{y_t\}$ does not follow but it will not be needed below.

For each ω let $A(\omega)$ be a measurable subset of $\boldsymbol{T}$. By FUBINI'S theorem and Theorem 5.1, there exists an Ω_0 with $\boldsymbol{P}(\Omega_0)=1$ such that if $\omega\in\Omega_0$, then (a) each $S_i(\omega)$, $i\in\bar{\boldsymbol{I}}$, is measurable (or Borel measurable if $\{x_t\}$ is); (b) $\mu[S_\infty(\omega)]=0$. The Lebesgue integral of $|y(\cdot,\omega)|$ then exists and

$$\int_{A(\omega)}|y(t,\omega)|\,dt=\sum_{i\in\boldsymbol{I}}|f(i)|\,\mu[A(\omega)\cap S_i(\omega)].$$

As usual we say that the sample function $y(\cdot,\omega)$ of the y-process is *summable* over $A(\omega)$ iff the above integral is finite. If $A(\omega)$ is an interval $(\tau(\omega),\tau'(\omega))$ we use the usual notation

$$\int_{\tau(\omega)}^{\tau'(\omega)} y(t,\omega)\,dt.$$

For given f and fixed $i,j\in\boldsymbol{I}$, $i\neq j$, we define

$$\mathscr{E}^{(i,j)}(f)=\sum_{k\in\boldsymbol{I}}e_k^{(i,j)}f(k);$$

and we say that $\mathscr{E}^{(i,j)}(f)$ exists iff $\mathscr{E}^{(i,j)}(|f|)<\infty$, i.e. the series above converges absolutely. It follows from the properties of $e_k^{(i,j)}$ that if $\mathscr{E}^{(i,j)}(f)$ exists for one pair (i,j) it exists for every pair. We define a sequence of entrance times inductively as follows. Let

$$\begin{aligned} \tau_1(\omega) &= \inf\{t\colon t\geqq 0;\ x(t,\omega)=i\}; \\ \sigma_n(\omega) &= \inf\{t\colon t>\tau_n(\omega);\ x(t,\omega)=j\}, \qquad n\geqq 2; \\ \tau_n(\omega) &= \inf\{t\colon t>\sigma_n(\omega);\ x(t,\omega)=i\}, \qquad n\geqq 1. \end{aligned}$$

These are all random variables and except for τ_1 which may equal zero with positive probability, they all have continuous distributions by Theorem 6.3. It is also easy to see that they are all optional relative to $\{x_t,\ t\in\boldsymbol{I}\}$; cf. Theorem 15.1 below. For a.a. ω, we have

$$(1) \qquad \lim_{n\to\infty}\sigma_n(\omega)=\lim_{n\to\infty}\tau_n(\omega)=\infty.$$

For suppose that the limit is $\zeta_\infty(\omega)<\infty$ for $\omega\in\Lambda$. Then if $\omega\in\Lambda$, as $t\uparrow\zeta_\infty(\omega)$, $x(t,\omega)$ has the two finite limiting values i and j. Hence $\boldsymbol{P}(\Lambda)=0$ by Theorem 6.1.

Theorem 1. If $\mathscr{E}^{(i,j)}(|f|)<\infty$, then the integrals

$$(2) \qquad \int_{\tau_n(\omega)}^{\tau_{n+1}(\omega)} y(t,\omega)\,dt, \quad \int_0^{\tau_1(\omega)} y(t,\omega)\,dt, \quad \int_0^A y(t,\omega)\,dt$$

exist and are finite for a.a. ω, for every $n\geqq 1$ and every constant $A\geqq 0$. Moreover, as functions of ω they are finite random variables.

Proof. We have

$$(3) \qquad \left\{ \begin{aligned} \int_{\tau_n(\omega)}^{\tau_{n+1}(\omega)} |y(t,\omega)|\,dt = \sum_{k\in\boldsymbol{I}} |f(k)|\,\{\mu[S_k(\omega)\cap(\tau_n(\omega),\sigma_n(\omega)]+ \\ \mu[S_k(\omega)\cap(\sigma_n(\omega),\tau_{n+1}(\omega))]\}. \end{aligned} \right.$$

Applying (11.17) to the post-τ_n process, we have

$$\boldsymbol{P}\{x(\tau_n(\omega)+u,\omega)\neq j,\ 0<u<t;\ x(\tau_n(\omega)+t,\omega)=k\}={}_jp_{ik}(t).$$

It follows by Fubini's theorem that

$$\boldsymbol{E}\{\mu[S_k(\omega)\cap(\tau_n(\omega),\sigma_n(\omega))]\}=\int_0^\infty {}_jp_{ik}(t)\,dt={}_jP_{ik}(\infty);$$

similarly

$$\boldsymbol{E}\{\mu[S_k(\omega)\cap(\sigma_n(\omega),\tau_{n+1}(\omega))]\}=\int_0^\infty {}_ip_{jk}(t)\,dt={}_iP_{jk}(\infty).$$

Substituting into (3), we obtain

$$\boldsymbol{E}\left\{\int_{\tau_n(\omega)}^{\tau_{n+1}(\omega)} |y(t,\omega)|\,dt\right\}=\sum_{k\in\boldsymbol{I}}|f(k)|\,e_k^{(i,j)}=\mathscr{E}^{(i,j)}(|f|)<\infty.$$

Hence by FUBINI's theorem again the first integral in (2) is a finite random variable. Next, for each $h \in \boldsymbol{I}$, if $\Delta_h = \{\omega\colon x_0(\omega) = h\}$, we have by a similar calculation

$$\boldsymbol{E}\left\{\int_0^{\tau_1(\omega)} |y(t, \omega)|\, dt \,\middle|\, \Delta_h\right\} \leqq \mathscr{E}^{(h,i)}(|f|) < \infty.$$

Hence for a.a. ω in Δ_h, the second integral in (2) is finite; this being true for every h we see that it is a finite random variable. Finally, given any A there exists a finite $n = n(A, \omega)$ such that $A < \tau_n(\omega)$, by (1). Hence we have

$$\int_0^A |y(t, \omega)|\, dt \leqq \int_0^{\tau_1(\omega)} |y(t, \omega)|\, dt + \sum_{\nu=1}^{n-1} \int_{\tau_\nu(\omega)}^{\tau_{\nu+1}(\omega)} |y(t, \omega)|\, dt.$$

The finiteness of the left member above now follows from what we have already proved. ▯

For each $\alpha = \tau_n$ or σ_n let us recall the pre-α and post-α fields defined in §§ 8—9. The random variable $x(\alpha(\cdot), \cdot)$ being a constant i or j it is measurable with respect to both $\mathscr{F}_\alpha$ and $\mathscr{F}'_\alpha$. We note also that the Borel field generated by the random variables of a stochastic process does not depend on the version of the process, since the field is augmented. By Theorem 9.5, the field $\mathscr{F}'_{\tau_n} \cap \mathscr{F}_{\tau_{n+1}}$ is independent of both $\mathscr{F}_{\tau_n}$ and $\mathscr{F}'_{\tau_{n+1}}$.

Theorem 2. Under the hypothesis of Theorem 1, the random variable

$$Y_n(\omega) = \int_{\tau_n(\omega)}^{\tau_{n+1}(\omega)} y(t, \omega)\, dt$$

is measurable $\mathscr{F}'_{\tau_n} \cap \mathscr{F}_{\tau_{n+1}}$. The sequence $\{Y_n,\ n \geqq 1\}$ is a sequence of independent and identically distributed random variables.

Proof. Let $\hat{y}(t, \omega) = y(t, \omega)$ if $\tau_n(\omega) \leqq t < \tau_{n+1}(\omega)$ and $= 0$ otherwise. Then $|\hat{y}(t, \omega)|$ is integrable over $[0, \infty)$ by Theorem 1 and the integral is $Y_n(\omega)$ for a.a. ω. By the remark after the Lemma in the proof of Theorem 5.3, the Lebesgue integral $\int_0^A \hat{y}(t, \cdot)\, dt$ for a finite A is the limit in probability of a sequence of suitably chosen Riemann sums. Now for each t, $\hat{y}(t, \cdot)$ is measurable $\mathscr{F}'_{\tau_n} \cap \mathscr{F}_{\tau_{n+1}}$ and so is each such Riemann sum. It follows that the Lebesgue integral is likewise measurable. This also proves the independence of the Y_n's by the remark preceding the theorem. Since the post-τ_n process and post-τ_{n+1} process have the same transition matrix, it follows from the Lemma in the proof of Theorem 5.3 that Y_n and Y_{n+1} have the same distribution. ▯

From here on the development of the theory of the functional process, in particular the various limit theorems whose discrete parameter analogues are given in detail in Part I, will be left as a long exercise for the interested reader. One theorem, however, which belongs in this category and whose discrete parameter analogue is too trivial to be given in Part I, will be inserted below for a later application (Example 20.3). It also illustrates an important case of the strong Markov property.

For $i \in \boldsymbol{I}$ and $s \geqq 0$ we define

$$\beta_i(s, \omega) = \mu[S_i(\omega) \cap (0, s)];$$

for each ω, $\beta_i(s, \omega)$ is nondecreasing and continuous in s. Let

$$\gamma_i(s, \omega) = \sup\{r\colon r \geqq 0;\ \beta_i(r, \omega) \leqq s\} = \inf\{r\colon r \geqq 0;\ \beta_i(r, \omega) > s\}$$

so that $\gamma_i(\cdot, \omega)$ is the right-continuous inverse function to $\beta_i(\cdot, \omega)$. In words, $\beta_i(s, \cdot)$ is the total amount of time spent in the state i up to the time s and $\gamma_i(s, \cdot)$ is the last time before the total amount of time spent in i exceeds s. For each ω, $\gamma_i(s, \omega)$ is strictly increasing with s in $s \in (0, \mu[S_i(\omega)])$ and we have for a.a. ω,

$$\lim_{t \downarrow \gamma_i(s, \omega)} x(t, \omega) = i \tag{4}$$

for each s in the above range by Theorem 6.1. Furthermore, for each $t \geqq 0$ we have

$$\{\omega\colon \gamma_i(s, \omega) < t\} = \{\omega\colon \beta_i(t, \omega) > s\} = \{\omega\colon \mu[S_i(\omega) \cap (0, t)] > s\}$$

which clearly belongs to $\mathscr{F}_t^0$; hence $\gamma_i(s) = \gamma_i(s, \cdot)$ is optional relative to $\{x_t,\ t \in \boldsymbol{T}\}$ for each $s \geqq 0$.

Using the notation of § 11 we define on Δ_i:

$${}_H c_{ij}(t, \omega) = \begin{cases} 1 & \text{if } x(t, \omega) = j \text{ and } x(s, \omega) \notin H \text{ for } \min(t, \varrho_i(\omega)) < s < t; \\ 0 & \text{otherwise;} \end{cases}$$

and

$${}_H C_{ij}(\omega) = \int_0^\infty {}_H c_{ij}(t, \omega)\, dt.$$

Thus starting from i, ${}_H C_{ij}$ is the total amount of time spent in the state j before entering the taboo set H unless $i \in H$, in which case the taboo is imposed only after the first exit from i; in symbols:

$${}_H C_{ij}(\omega) = \mu[S_i(\omega) \cap (0, \inf_{k \in H} \alpha_{ik}(\omega))].$$

Clearly ${}_H C_{ij}$ is a random varbiale; its distrbiution is given by the following theorem.

Theorem 3. We have for all $t \geqq 0$,

$$(5) \qquad \boldsymbol{P}\{{}_HC_{ij}(\omega) \leqq t \mid \Delta_i\} = 1 - {}_H\tilde{F}_{ij}(\infty)\, e^{-{}_HP_{jj}^{-1}(\infty)t}$$

where ${}_H\tilde{F}_{ij}(\infty) = {}_HF_{ij}(\infty)$ if $i \neq j$ and $=1$ if $i = j$.

Proof. It is sufficient to consider the case $i = j$. Using the random variable $\gamma_i(s)$ defined above, we have

$$\Lambda = \Delta_i \cap \{\omega : {}_HC_{ii}(\omega) > s\} = \Delta_i \cap \{\omega : \gamma_i(s, \omega) < \inf_{k \in H} \alpha_{ik}(\omega)\} \in \mathscr{F}_{\gamma_i(s)}$$

where $\mathscr{F}_{\gamma_i(s)}$ is the pre-$\gamma_i(s)$ field. Let $s \geqq 0$, $s' \geqq 0$; because of (4) we can apply Theorem 9.5 and Theorem 11.6 to the post-$\gamma_i(s)$ process:

$$\begin{aligned} \boldsymbol{P}\{{}_HC_{ii}(\omega) > s + s' \mid \Delta_i\} &= \boldsymbol{P}\{\Lambda ; \mu[S_i(\omega) \cap (\gamma_i(s, \omega), \inf_{k \in H} \alpha_{ik}(\omega))] > s' \mid \Delta_i\} \\ &= \boldsymbol{P}\{\Lambda \mid \Delta_i\}\, \boldsymbol{P}\{{}_HC_{ii}(\omega) > s' \mid \Delta_i\} = \boldsymbol{P}\{{}_HC_{ii}(\omega) > s \mid \Delta_i\}\, \boldsymbol{P}\{{}_HC_{ii}(\omega) > s' \mid \Delta_i\}. \end{aligned}$$

It follows from the well known theorem on HAMEL'S equation that there is a constant c such that

$$\boldsymbol{P}\{{}_HC_{ii}(\omega) > s \mid \Delta_i\} = e^{-cs}.$$

To determine c we observe that it is equal to the reciprocal of the mean

$$\boldsymbol{E}\{{}_HC_{ii} \mid \Delta_i\} = \int_0^\infty \boldsymbol{E}\{{}_Hc_{ii}(t, \cdot) \mid \Delta_i\}\, dt = \int_0^\infty {}_Hp_{ii}(t)\, dt = {}_HP_{ii}(\infty).$$

The last two equations yield (5) in case $i = j$. ▯

We indicate an alternative proof of Theorem 3 in case ${}_HP_{jj}(\infty) < \infty$, by calculating the moments. As in the above we have $\boldsymbol{E}\{{}_HC_{ij} \mid \Delta_i\} = {}_HP_{ij}(\infty)$. For the second moment we have

$$\begin{aligned} \boldsymbol{E}\{{}_HC_{ij}^2 \mid \Delta_i\} &= \int_0^\infty \int_0^t {}_Hp_{ij}(s)\, {}_Hp_{jj}(t-s)\, ds\, dt + \int_0^\infty \int_t^\infty {}_Hp_{ij}(t)\, {}_Hp_{jj}(s-t)\, ds\, dt \\ &= 2\, {}_HP_{ij}(\infty)\, {}_HP_{jj}(\infty). \end{aligned}$$

In general it is easily seen that the n-th moment is

$$\boldsymbol{E}\{{}_HC_{ij}^n \mid \Delta_i\} = n!\, {}_HP_{ij}(\infty)\, {}_HP_{jj}^{n-1}(\infty).$$

Recalling that ${}_HP_{ij}(\infty) = {}_HF_{ij}(\infty)\, {}_HP_{jj}(\infty)$ this yields another proof of Theorem 3.

In case j is stable a more specific result can be given. We may suppose that $j \notin H$ for otherwise (5) is trivial. Define ${}_HN_{ij}(\omega)$ on Δ_i to be the number of j-intervals preceding $\inf_{k \in H} \alpha_{ij}(\omega)$. Clearly we have if $i \neq j$,

$$\boldsymbol{P}\{{}_HN_{ij}(\omega) = 0 \mid \Delta_i\} = 1 - {}_HF_{ij}(\infty);$$

$$\boldsymbol{P}\{{}_HN_{ij}(\omega) = n+1 \mid \Delta_i\} = {}_HF_{ij}(\infty)\, {}_HF_{jj}^n(\infty)\, [1 - {}_HF_{jj}(\infty)\}, \qquad n \geqq 0.$$

Let the length of the n-th j-interval of $x(\cdot, \omega)$ be $\varrho_n(\omega)$. It follows from Theorem 15.2 in the next section that

$$\boldsymbol{P}\Big\{{}_H N_{ij}(\omega) = n+1;\ \sum_{\nu=1}^{n+1} \varrho_n(\omega) \leqq t \,|\, \varDelta_i\Big\}$$

$$= \boldsymbol{P}\{{}_H N_{ij}(\omega) = n+1 \,|\, \varDelta_i\} \int_0^t \frac{(q_j s)^n}{n!}\, q_j\, e^{-q_j s}\, ds$$

and consequently

$$\boldsymbol{P}\{{}_H C_{ij}(\omega) \leqq t \,|\, \varDelta_i\} = 1 - {}_H F_{ij}(\infty) + {}_H F_{ij}(\infty)\left[1 - e^{-q_j(1 - {}_H F_{jj}(\infty))t}\right].$$

Now $q_j(1 - {}_H F_{jj}(\infty)) = {}_H P_{jj}^{-1}(\infty)$ as follows easily from (11.19) by taking $i = j = k$ and integrating over $[0, \infty)$; cf. (13.16). Hence the above formula reduces to (5); the case $i = j$ is similar.

Notes. Some, though not all, details have been given for the proofs of Theorems 1 and 2 to illustrate the delicate points involved in extending a discrete parameter result to a "perfectly obvious" analogue in the continuous parameter case.

LÉVY [1], [2], [3] has the idea of using $\beta_i(s)$ as a "new time parameter". More generally we may use the accumulative time spent in a chosen set of states as the new time parameter in a new process derived from the original one by "annihilating" the complementary set of states. This is a case of *optional sampling* which is not discussed here. It is related to but different from considering taboo probabilities in which the time parameter is unchanged. LÉVY used this device to deduce several previous results which are proved here by other means.

§ 15. Post-exit process

One of the most interesting case of the strong Markov property which has been studied so far is that where the optional random variable is an exit time from a stable state. In contrast to the case of an entrance time, which can be defined also for an instantaneous state, we have here already the general situation where the adjoined state ∞ may appear at the optional time. It turns out however that there is still a certain independence between the past and the future.

The M.C. $\{x_t, t \in \boldsymbol{T}\}$ is taken to be the x_+ version. Let i be a stable but not absorbing state, namely $0 < q_i < \infty$; α with domain of finiteness $\varDelta$ and optional relative to $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$ such that $x(\alpha(\omega), \omega) = i$ for all $\omega \in \varDelta$; $\{\xi_t, t \in \boldsymbol{T}\}$ the post-α process where $\xi(t, \omega) = x(\alpha(\omega) + t, \omega)$. According to Theorems 9.3 and 9.5 the ξ-process is a M.C. with $\{\mathscr{F}_{a+t}, t \in \boldsymbol{T}\}$ as an admissible family and with $\boldsymbol{P}\{\xi_0(\omega) = i\} = 1$; and the Borel fields $\mathscr{F}_\alpha$ and $\mathscr{F}'_\alpha$ are independent. We now define on $\varDelta$:

$$\beta(\omega) = \inf\{t\colon t > \alpha(\omega),\ x(t,\omega) \neq i\};$$

$$\varrho(\omega) = \beta(\omega) - \alpha(\omega).$$

It follows that

$$\varrho(\omega) = \inf\{t\colon t > 0;\ x(\alpha(\omega)+t,\omega) \neq i\} = \inf\{t\colon t > 0;\ \xi(t,\omega) \neq i\}$$

and consequently for each $s > 0$,

$$\varDelta \cap \{\omega\colon \varrho(\omega) > s\} = \varDelta \cap \{\omega\colon \xi(t,\omega) \equiv i,\ 0 \leqq t \leqq s\}. \tag{1}$$

By Corollary 3 to Theorem 9.3, $\{\xi_t\}$ is separable so that the probability of the right member above is defined. We infer from (1) that ϱ is optional relative to $\{\xi_t, \mathscr{F}_{\alpha+t}, t \in \boldsymbol{T}\}$. From this we infer that β is optional relative to $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$ by the following general result.

Theorem 1. If α is optional relative to $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$ and ϱ is optional relative to $\{\xi_t, \mathscr{F}_{\alpha+t}, t \in \boldsymbol{T}\}$ where $\{\xi_t, t \in \boldsymbol{T}\}$ is the post-α process; then $\alpha + \varrho$ is optional relative to $\{x_t, \mathscr{F}_t, t \in \boldsymbol{T}\}$.

Proof. Let $\mathscr{G}_s = \mathscr{F}\{\xi_t, 0 \leqq t \leqq s\}$. We have

$$\{\omega\colon \alpha(\omega) + \varrho(\omega) < t\} = \bigcup_{r<t} \{\omega\colon \alpha(\omega) < r;\ \varrho(\omega) < t - r\}, \tag{2}$$

where r runs through the rationals. Since $\{\omega\colon \varrho(\omega) < t-r\} \in \mathscr{F}_{\alpha+t-r}$ we have by the definition of $\mathscr{F}_{\alpha+t-r}$,

$$\{\omega\colon \alpha(\omega) < r;\ \varrho(\omega) < t-r\} \in \mathscr{F}_{r+t-r} = \mathscr{F}_t.$$

Hence the union on the right side of (2) also belongs to $\mathscr{F}_t$ and $\alpha + \varrho$ is optional as asserted. ▯

Remark. If α and β where $\beta \geqq \alpha$ are both optional relative to $\{x_t\}$, it does not follow that $\beta - \alpha$ is optional relative to $\{\xi_t\}$. But β is optional relative to $\{\mathscr{F}_{\alpha+t}, t \in \boldsymbol{T}\}$; see CHUNG and DOOB [1]. Example: $\beta = \max(\alpha, c)$ where c is a positive constant.

Let us denote the post-β process by $\{\eta_t, t \in \boldsymbol{T}\}$ where

$$\eta_t(\omega) = \eta(t,\omega) = x(\beta(\omega)+t,\omega) = \xi(\varrho(\omega)+t,\omega);$$

and let $\mathscr{F}'_\beta$ be the post-β field. The stochastic process $\{\eta_t, t \in \boldsymbol{T}\}$ is called a *post-exit process from the stable state i*.

The following theorem, basic for a post-exit process, asserts that the three Borel fields $\mathscr{F}_\alpha$, $\mathscr{F}\{\varrho\}$, $\mathscr{F}'_\beta$ are independent, where $\mathscr{F}\{\varrho\}$ is the Borel field generated by the single random variable ϱ; note that all these fields contain only subsets of $\varDelta$.

Theorem 2. Let $\varLambda \in \mathscr{F}_\alpha$, $\mathsf{M} \in \mathscr{F}'_\beta$ and $s \geqq 0$, then

$$\boldsymbol{P}\{\varLambda;\ \varrho(\omega) > s;\ \mathsf{M}\} = \boldsymbol{P}(\varLambda)\, e^{-q_i s}\, \boldsymbol{P}\{\mathsf{M} \mid \varDelta\}. \tag{3}$$

Remark. This does not imply that $\mathscr{F}_\beta$ and $\mathscr{F}'_\beta$ are independent since $\mathscr{F}_\beta$ is in general larger than the smallest Borel field containing $\mathscr{F}_\alpha$ and $\mathscr{F}\{\varrho\}$; the random variable x_β being $\mathscr{F}_\beta$-measurable in the x_+-version.

Proof. By Theorems 9.5, we have

$$\boldsymbol{P}\{\Lambda;\varrho(\omega)>s;\mathsf{M}\,|\,\Delta\}=\boldsymbol{P}\{\Lambda\,|\,\Delta\}\boldsymbol{P}\{\varrho(\omega)>s;\mathsf{M}\,|\,\Delta\}.$$

Hence it remains to prove that

$$\boldsymbol{P}\{\varrho(\omega)>s;\mathsf{M}\,|\,\Delta\}=e^{-q_i s}\boldsymbol{P}\{\mathsf{M}\,|\,\Delta\}. \tag{4}$$

It is sufficient to prove (4) when

$$\mathsf{M}=\Delta\cap\{\omega:\eta(t_\nu,\omega)=j_\nu,\ 1\leqq\nu\leqq N\}$$

for arbitrary $0\leqq t_1<\cdots<t_N$ and $j_1,\ldots,j_N\in\boldsymbol{I}$. For $s\geqq 0$ let $\Delta^{(s)}=\{\omega:\xi(s,\omega)=i\}$ and define on $\Delta^{(s)}$:

$$\varrho^{(s)}(\omega)=\inf\{t:t>s;\ \xi(t,\omega)\neq i\};\qquad \beta^{(s)}(\omega)=\alpha(\omega)+\varrho^{(s)}(\omega);$$
$$\eta^{(s)}(t,\omega)=\xi\big(\varrho^{(s)}(\omega)+t,\omega\big)=x\big(\beta^{(s)}(\omega)+t,\omega\big);$$
$$\mathsf{M}^{(s)}=\Delta^{(s)}\cap\{\omega:\eta^{(s)}(t_\nu,\omega)=j_\nu,\ 1\leqq\nu\leqq N\}.$$

Thus for $s=0$ the notation reduces to the previous one without subscript. Observe that if $\varrho(\omega)>s$, then $\varrho(\omega)=\varrho^{(s)}(\omega)$ and $\mathsf{M}=\mathsf{M}^{(s)}$, hence

$$\begin{aligned}\boldsymbol{P}\{\varrho(\omega)>s;\mathsf{M}\,|\,\Delta\}&=\boldsymbol{P}\{\varrho(\omega)>s;\mathsf{M}^{(s)}\,|\,\Delta\}\\&=\boldsymbol{P}\{\varrho(\omega)>s\,|\,\Delta\}\boldsymbol{P}\{\mathsf{M}^{(s)}\,|\,\Delta;\varrho(\omega)>s\}\\&=\boldsymbol{P}\{\varrho(\omega)>s\,|\,\Delta\}\boldsymbol{P}\{\mathsf{M}^{(s)}\,|\,\Delta;\xi(s,\omega)=i\}\\&=\boldsymbol{P}\{\varrho(\omega)>s\,|\,\Delta\}\boldsymbol{P}\{\mathsf{M}^{(s)}\,|\,\Delta^{(s)}\}.\end{aligned}$$

In the equation before the last we have used the fact that $\{\omega:\xi(s,\omega)=i\}\supset\{\omega:\varrho(\omega)>s\}\in\mathscr{F}_{\alpha+s}$ and that the family $\{\mathscr{F}_{\alpha+t},t\in\boldsymbol{T}\}$ is admissible relative to $\{\xi_t,t\in\boldsymbol{T}\}$.

By Theorem 11.7, taking H to be $\boldsymbol{I}-\{\mathrm{i}\}$ in (11.17), we have

$$\boldsymbol{P}\{\varrho(\omega)>s\,|\,\Delta\}=\boldsymbol{P}\{\xi(u,\omega)\equiv i,\ 0\leqq u\leqq s\,|\,\Delta\}={}_ip_{ii}(s)=e^{-q_i s}. \tag{5}$$

To conclude the proof of Theorem 2 it remains to prove that

$$\boldsymbol{P}\{\mathsf{M}\,|\,\Delta\}=\boldsymbol{P}\{\mathsf{M}^{(s)}\,|\,\Delta^{(s)}\}. \tag{6}$$

It may appear that (6) is an immediate consequence of the stationarity of the transition probabilities of $\{x_t\}$. To render this thought faultless we recall the proof of Theorem 9.2 and observe that

$$\mathsf{M}^{(s)} \doteq \varDelta^{(s)} \cap \limsup_{n} \bigcup_{m=0}^{\infty} \Big\{\omega\colon s+\frac{m}{n} \leqq \varrho^{(s)}(\omega) < s+\frac{m+1}{n};$$
$$\xi\left(s+\frac{m+1}{n}+t_\nu, \omega\right) = j_\nu, 0 \leqq \nu \leqq N\Big\}.$$

It now follows that the probability on the right side of (6) does not depend on s. ▯

Corollary 1. If K (in $\varDelta$) belongs to the smallest Borel field containing both $\mathscr{F}_\alpha$ and $\mathscr{F}'_\beta$, then

$$\boldsymbol{P}\{\mathsf{K}; \varrho(\omega) > s\} = \boldsymbol{P}(\mathsf{K})\, e^{-q_i s}.$$

As an important application of Theorem 2, let i be a stable state and

$$\tau_1(i;\omega) = \inf\{t\colon t>0;\ x(t,\omega)=i\}$$

and let $\varDelta_1(i)$ be the ω-set on which the t-set between the braces is nonempty, namely $\varDelta_1(i)$ is the domain of finiteness of $\tau_1(i;\omega)$. Let

$$\tau'_1(i;\omega) = \inf\{t\colon t>\tau_1(i;\omega);\ x(t,\omega) \neq i\};$$

since i is stable $\tau'_1(t,\omega)$ is finite a.e. on $\varDelta_1(i)$. Assuming that $\tau_n(i;\omega)$ and $\tau'_n(i;\omega)$ have been defined with the domain of definition and finiteness $\varDelta_n(i)$, we define inductively

$$\tau_{n+1}(i;\omega) = \inf\{t\colon t>\tau'_n(i;\omega);\ x(t,\omega)=i\};$$

$\varDelta_{n+1}(i)$ to be the ω-set on which the above t-set is nonempty; and

$$\tau'_{n+1}(i;\omega) = \inf\{t\colon t>\tau_{n+1}(i;\omega);\ x(t,\omega) \neq i\}.$$

It is easy to see that each $\tau_n(i)$ or $\tau'_n(i)$ is a random variable which is optional relative to $\{x_t, \mathscr{F}_t^0, t\in\boldsymbol{T}\}$. Let

$$\lambda_n(i;\omega) = \tau'_n(i;\omega) - \tau_n(i;\omega).$$

We call $\tau_n(i)$, $\tau'_n(i)$ and $\lambda_n(i)$ respectively *the n-th entrance, exit, and sojourn time of the stable state i.*

We can now state the following important corollary to Theorem 2.

Corollary 2. Let $\{i_\nu, 1\leqq\nu\leqq N\}$ be stable states and let $\{n_\nu, 1\leqq\nu\leqq N\}$ be positive integers such that if $\nu\neq\nu'$ then either $i_\nu \neq i_{\nu'}$ or $n_\nu \neq n_{\nu'}$. Let $\mathscr{G}_n(j)$ be the smallest field containing $\mathscr{F}_{\tau_n(j)}$ and $\mathscr{F}'_{\tau'_n(j)}$. If $\mathsf{K} \in \bigcap_{\nu=0}^{n} \mathscr{G}_{n_\nu}(i_\nu)$ then

$$\boldsymbol{P}\{\mathsf{K}; \varrho_{n_\nu}(i_\nu, \omega) > s_\nu, 1\leqq\nu\leqq N\} = \boldsymbol{P}(\mathsf{K}) \prod_{\nu=1}^{N} e^{-q_{i_\nu} s_\nu}$$

In particular the random variables $\lambda_{n_\nu}(i_\nu)$, $1\leqq\nu\leqq N$, are independent relative to $\bigcap\limits_{\nu=1}^{N}\Delta_{n_\nu}(i_\nu)$, namely under the hypothesis that they are all finite.

As an application of Corollary 2, we give a proof of a special case of Theorem 1.6 or 5.2 which is instructive.

Theorem 3. Suppose that $q_k<\infty$ for every $k\in\boldsymbol{I}$. Then for each i and j, the function p_{ij} is either identically zero or never zero.

Proof. Even without the additional hypothesis, we know from Theorem 1.5 that in a standard transition matrix p_{ii} is never zero. Hence we may suppose thgt $i\neq j$; we may also suppose $\boldsymbol{P}\{x_0(\omega)=i\}=1$ without loss of generality. Recalling (11.13), since $\alpha_{ij}=\tau_1(j)$ under Δ_i, we have $F_{ij}(t)=\boldsymbol{P}\{\tau_1(j;\omega)\leqq t\}$. Clearly by (13.3) it is sufficient to prove that if $F_{ij}(t)>0$ for some $t>0$, then $F_{ij}(t)>0$ for an arbitrarily small $t>0$. All statements below concerning ω will be valid for a.a. ω. Since $\mu[S_\infty(\omega)]=0$ by Theorem 5.1 we have

$$(7)\qquad \tau_1(j;\omega)=\sum_{k\in\boldsymbol{I}}\mu[S_k(\omega)\cap(0,\tau_1(j,\omega))].$$

Let $\varepsilon_0>0$ be given; by hypothesis there exists an ε, $0<\varepsilon<\varepsilon_0$, and an ω-set Λ_1 with $\boldsymbol{P}(\Lambda_1)>0$ such that if $\omega\in\Lambda_1$ then $\varepsilon<\tau_1(j,\omega)<\infty$. The series in (7) therefore converges on Λ_1; it follows that there exists a finite set K with N elements and $\Lambda_2\subset\Lambda_1$ with $\boldsymbol{P}(\Lambda_2)>0$ such that if $\omega\in\Lambda_2$,

$$\sum_{k\in K}\mu[S_k(\omega)\cap(0,\tau_1(j;\omega))]>\tau_1(j;\omega)-\frac{\varepsilon}{4}.$$

(This as well as the next statement is a simple case of EGOROV's theorem.) Since all states are stable each $S_k(\omega)$ is the union of intervals of lengths $\lambda_n(k;\omega)$, $n\geqq 1$, it follows again that there exists an $N'<\infty$ and $\Lambda_3\subset\Lambda_2$ with $\boldsymbol{P}(\Lambda_3)>0$ such that if $\omega\in\Lambda_3$, there exist $\{k_\nu,\ n_\nu,\ 1\leqq\nu\leqq m\}$ with $k_\nu\in K$ and $1\leqq n_\nu\leqq N'$ such that

$$\text{(a)}\qquad \tau_{n_1}(k_1;\omega)<\cdots<\tau_{n_m}(k_m;\omega)<\tau_1(j,;\omega)$$

and

$$\text{(b)}\qquad \tau_1(j;\omega)-\frac{\varepsilon}{2}<\sum_{\nu=1}^{m}\lambda_{n_\nu}(k_\nu;\omega)\leqq\tau_1(j;\omega).$$

Consider all possible ordering of $\{\tau_n(k),\ k\in K,\ 1\leqq n\leqq N'\}$ subject to the restriction that $\tau_n(k)<\tau_{n'}(k)$ if $n<n'$; and consider all the initial sections of these ordered sequences, a typical one of which may be denoted by $\varphi=\{\tau_{n_1}(k_1)<\cdots<\tau_{n_m}(k_m)\}$. Let $\Lambda_{3\varphi}$ denote the subset of Λ_3 for which (a) and (b) hold. Since the number of φ's is finite there exists at least one φ for which $\boldsymbol{P}(\Lambda_{3\varphi})>0$. Without complicating the notation let this be the φ indicated above. For $\omega \in \Lambda_{3\varphi}$ we put

$$\delta_l(\omega) = \tau_{n_{l+1}}(k_{l+1};\omega) - \tau'_{n_l}(k_l;\omega), \qquad 1 \leqq l \leqq m-1;$$

$$\delta_m(\omega) = \tau_1(j;\omega) - \tau_{n_m}(k_m;\omega).$$

Furthermore we put

$$\Lambda_4 = \left\{\omega\colon \tau_{n_1}(k_1;\omega) < \cdots < \tau_{n_m}(k_m;\omega) < \tau_1(j;\omega) < \infty;\ \sum_{l=1}^{m} \delta_l(\omega) < \frac{\varepsilon}{2}\right\}.$$

It follows from the definition of $\Lambda_{3\varphi}$ that $\boldsymbol{P}(\Lambda_4) \geqq \boldsymbol{P}(\Lambda_{3\varphi}) > 0$. On the other hand, applying Corollary 2 to Theorem 2, we have

$$\begin{aligned}\boldsymbol{P}\{\tau_1(j;\omega) < \varepsilon\} &\geqq \boldsymbol{P}\left\{\Lambda_4;\ \lambda_{n_\nu}(k_\nu,\omega) < \frac{\varepsilon}{2m},\ 1 \leqq \nu \leqq m\right\}\\ &= \boldsymbol{P}(\Lambda_4) \prod_{\nu=1}^{m} e^{-q_{k_\nu}\varepsilon/2m} > 0.\end{aligned}$$

This concludes the proof since ε_0 is arbitrary. ▯

From now on in this section we shall deal with a fixed post-exit process. There is then no loss of generality if we assume $\alpha \equiv 0$ in the above so that $\boldsymbol{P}\{x_0(\omega) = i\} = 1$. We have thus

$$\varrho_i(\omega) = \inf\{t\colon t > 0;\ x(t,\omega) \neq i\}; \qquad \eta(t,\omega) = x(\varrho_i(\omega) + t, \omega)$$

defined a.e. on $\Delta = \Omega$.

Theorem 4. For each $j \in \boldsymbol{I}$ there is a function r_{ij} continuous in $[0, \infty)$ such that

$$(8) \qquad \boldsymbol{P}\{x(t,\omega) = j \mid \varrho_i = s\} = r_{ij}(t-s)$$

for each $t \geqq 0$ and a.a. (Lebesgue measure) $s \in [0, t]$. Moreover we have

$$(9) \qquad \boldsymbol{P}\{\eta(t,\omega) = j\} = r_{ij}(t).$$

Proof. It follows from Theorem 2 that the random variable ϱ_i and the Borel field $\mathscr{F}\{\eta_t, t \in \boldsymbol{I}\}$ are independent. Hence Corollary to Theorem 9.4 is applicable and the theorem follows. Note that here the $A(\Delta;\cdot)$ measure is that determined by the distribution function $1 - e^{-q_i t}$. ▯

The analytical content of (8) is recapitulated in the next theorem, which also serves to identify r_{ij} analytically.

Theorem 5. Let i be stable and $j \in \boldsymbol{I}$. We have for every $t \geqq 0$,

$$(10) \qquad p_{ij}(t) = \delta_{ij}\, e^{-q_i t} + \int_0^t q_i\, e^{-q_i(t-s)}\, r_{ij}(s)\, ds,$$

where r_{ij} is continuous in $[0, \infty)$. Consequently $p'_{ij}(t)$ exists in $[0, \infty)$ and

$$(11) \qquad p'_{ij}(t) = q_i[r_{ij}(t) - p_{ij}(t)].$$

Proof. We may suppose that $q_i > 0$; otherwise the assertions are trivial. In view of (8), the equation (10) is simply the analytic expression of the relation

$$\begin{aligned} p_{ij}(t) &= \boldsymbol{P}\{\varrho_i(\omega) > t;\ x(t,\omega) = j\} + \boldsymbol{P}\{\varrho_i(\omega) \leqq t;\ x(t,\omega) = j\} \\ &= \boldsymbol{P}\{\varrho_i(\omega) > t;\ x(t,\omega) = j\} + \int\limits_{\{\omega:\ \varrho_i(\omega) \leqq t\}} \boldsymbol{P}\{x(t,\omega) = j \mid \varrho_i\} \boldsymbol{P}(d\omega). \end{aligned}$$

Equation (11) follows from (10) upon differentiation since r_{ij} is continuous. ▯

We have thus a new and probabilistic proof of the main assertions of Theorem 3.1, namely the existence of the continuous derivative p'_{ij}. On the other hand, the latter can be used to prove the probabilistic statement (8) via (10). Similarly (3.3) and (3.5) have their counterparts in the following equations:

$$\sum_j r_{ij}(t) = 1; \tag{12}$$

$$r_{ij}(s+t) = \sum_k r_{ik}(s)\, p_{kj}(t). \tag{13}$$

These are particular cases of (8.12) and (8.13) and express the fundamental Markov property of the post-exit process. Theorem 2.6 is also given a new meaning as follows.

Theorem 6. We have for every $j \in \boldsymbol{I}$,

$$\boldsymbol{P}\{\eta(0,\omega) = j\} = r_{ij}(0) = \frac{(1-\delta_{ij})\, q_{ij}}{q_i}; \tag{14}$$

and

$$\boldsymbol{P}\{\eta(0,\omega) = \infty\} = 1 - \frac{\sum\limits_{j \neq i} q_{ij}}{q_i}.$$

Proof. This is the particular case of Theorem 9.1. The evaluation of $r_{ij}(0)$ follows from (11) and Theorems 2.4 and 2.5. ▯

Theorem 6 is illustrative of the general sample function behavior described in Theorem 7.4. According as j is stable or instantaneous, as $s \downarrow \varrho_i(\omega)$, $r_{ij}(0)$ is the probability of case (a) or (b) (read j for the i there), and $1 - \sum_j r_{ij}(0)$ is the probability of case (c) of that theorem. Clearly $r_{ii}(0) = 0$.

We proceed to derive some further relations. Let $\gamma_{ij} = \alpha_{ij} - \varrho_i$; then by Theorem 2 the two random variables ϱ_i and γ_{ij} are independent. We have

$$\gamma_{ij}(\omega) = \inf\{t\colon t \geqq 0\colon \eta(t,\omega) = j\}.$$

Let R_{ij} be the distribution function of γ_{ij}:

$$R_{ij}(t) = \boldsymbol{P}\{\gamma_{ij}(\omega) \leqq t\}.$$

Theorem 7. R_{ij} is continuous in $[0, \infty)$ and

$$F_{ij}(t)=\int_0^t [1-e^{-q_i(t-s)}]\, dR_{ij}(s); \tag{15}$$

$$r_{ij}(t)=\int_0^t p_{jj}(t-s)\, dR_{ij}(s). \tag{16}$$

Furthermore F'_{ij} exists and is continuous in $[0, \infty)$ and

$$F'_{ij}(t)=q_i[R_{ij}(t)-F_{ij}(t)]. \tag{17}$$

F''_{ij} and R'_{ij} exist a.e.; if j is stable then p''_{ij} and r'_{ij} exist a.e.

Proof. The continuity of R_{ij} follows from the definition of γ_{ij} and Theorem 6.3; (15) follows from the independence of ϱ_i and γ_{ij}. Formula (16) is the analogue of formula (13.3) for the η-process. Integrating by parts in (15) we obtain

$$F_{ij}(t)=\int_0^t q_i\, e^{-q_i(t-s)} R_{ij}(s)\, ds. \tag{18}$$

Hence F_{ij} has a continuous derivative given by (17). Since the right member of (17) is a function of bounded variation, F''_{ij} exists a.e. and so does R'_{ij} by (17). By (12.21), $\int_0^t |p'_{jj}(s)|\, ds<\infty$; hence we may differentiate (13.3) to obtain

$$p'_{ij}(t)=-\delta_{ij}\, q_i\, e^{-q_i t}+F'_{ij}(t)+\int_0^t p'_{jj}(t-s)\, F'_{ij}(s)\, ds.$$

If j is stable, $p'_{jj}(t)$ is bounded near $t=0$ by Theorem 3.1. It follows from the above that p'_{jj} exists a.e; and so does r'_{ij} by (11). ▯

Theorem 8. We have

$$F_{ij}(\infty)=R_{ij}(\infty); \qquad \lim_{t\to\infty} p_{ij}(t)=\operatorname*{llm}_{t\to\infty} r_{ij}(t); \qquad \lim_{t\to\infty} F'_{ij}(t)=\lim_{t\to\infty} p'_{ij}(t)=0.$$

Proof. The first relation follows from (18); the second from (16) and Theorem 10.1; the third now follows from (17) and (11). ▯

We add a brief discussion of taboo probabilities for the post-exit process $\{\eta_t\}$. Define

$${}_H r_{ij}(t)=\boldsymbol{P}\{\eta(s,\omega)\notin H,\ 0<s<t;\ \eta(t,\omega)=j\}. \tag{19}$$

Then we have if $s>0$, $t\geqq 0$, noting that ${}_H r_{ik}(s)\equiv 0$ if $k\in H$,

$${}_H r_{ij}(s+t)=\sum_k {}_H r_{ik}(s)\, {}_H p_{kj}(t).$$

It follows from this that ${}_H r_{ij}$ is a continuous function in $[0, \infty)$ and that the limit exists as $t\downarrow 0$. We have in analogy with (10), and using the definition of ${}_H p_{ij}$ in (11.4),

$$(20)\qquad {}_H p_{ij}(t) = \delta_{ij}\, e^{-q_i t} + \int_0^t q_i\, e^{-q_i(t-s)}\, {}_H r_{ij}(s)\, ds$$

so that ${}_H p_{ij}$ has a continuous derivative in $[0, \infty)$ and

$${}_H p'_{ij}(t) = q_i [{}_H r_{ij}(t) - {}_H p_{ij}(t)]$$

in analogy with (11), and so on. It follows that the Laplace transform of ${}_H p'_{ij}$ exists for $x > 0$, and

$$\int_0^\infty e^{-xt}\, {}_H p'_{ij}(t)\, dt = x\, {}_H \hat p_{ij}(x) - \delta_{ij}.$$

Hence if $\hat r(x) = \int_0^\infty e^{-xt} r(t)\, dt$ is the Laplace transform of r, we have from (20)

$$(21)\qquad (q_i + x)\, {}_H \hat p_{ij}(x) = \delta_{ij} + q_i\, {}_H \hat r_{ij}(x).$$

The relation (21) leads to an analogue of the last entrance decomposition formula (I.9.2), without making use of the results of § 12. Taking Laplace transforms of (11.19) we have, omitting the variable x wherever possible,

$$(22)\qquad {}_H \hat p_{ij} = {}_{k,H} \hat p_{ij} + {}_H \hat p_{kj}\, {}_H \hat F_{ik};$$

particular cases of which are

$${}_H \hat p_{ik} = \delta_{ik} (q_i + x)^{-1} + {}_H \hat p_{kk}\, {}_H \hat F_{ik};$$

$${}_H \hat p_{kj} = {}_{k,H} \hat p_{kj} + {}_H \hat p_{kj}\, {}_H \hat F_{kk}.$$

Substituting the last two formulas into (22) we obtain

$$\begin{aligned} {}_H \hat p_{ij} &= {}_{k,H} \hat p_{ij} + {}_H \hat p_{kj} ({}_H \hat p_{kk})^{-1} [{}_H \hat p_{ik} - \delta_{ik} (q_i + x)^{-1}] \\ &= {}_{k,H} \hat p_{ij} + {}_{k,H} \hat p_{kj} (1 - {}_H \hat F_{kk})^{-1} ({}_H \hat p_{kk})^{-1} [{}_H \hat p_{ik} - \delta_{ik} (q_i + x)^{-1}]. \end{aligned}$$

If k is also a stable state then

$${}_H \hat p_{kk} = (q_k + x)^{-1} + {}_H \hat p_{kk}\, {}_H \hat F_{kk}$$

so that the preceding formula reduces to

$$(23)\qquad {}_H \hat p_{ij} = {}_{k,H} \hat p_{ij} + (q_k + x)\, {}_{k,H} \hat p_{kj} [{}_H \hat p_{ik} - \delta_{ik} (q_i + x)^{-1}].$$

Substituting from (21) we obtain after some simplification

$$(1 - \delta_{kj})\, {}_H \hat p_{ij} = (1 - \delta_{ik})\, {}_{k,H} \hat p_{ij} + q_k\, {}_{k,H} \hat r_{kj}\, {}_H \hat p_{ik}.$$

The interesting case is where $i \neq k \neq j$; we have then

$$(24)\qquad {}_H p_{ij}(t) = {}_{k,H} p_{ij}(t) + q_k \int_0^t {}_H p_{ik}(t - s)\, {}_{k,H} r_{kj}(s)\, ds$$

where i and k are stable but j is arbitrary. This is the analogue of (I.9.2).

A particular case of (23) ,where H is empty and $i=k$, is

(25) $$\hat{p}_{ij}=(q_i+x)\,\hat{p}_{ii}\,{}_i\hat{p}_{ij};$$

Using (21) with $i=j$, we obtain

(26) $$\hat{p}_{ij}=(1+q_i\,\hat{r}_{ii})\,{}_i\hat{p}_{ij}$$

or

$$p_{ij}(t)={}_ip_{ij}(t)+\int_0^t q_i\,r_{ii}(s)\,{}_ip_{ij}(t-s)\,ds.$$

From (20) we deduce for $i\neq j$:

$${}_i\hat{p}_{ij}=q_i\,{}_i\hat{r}_{ij}(q_i+x)^{-1}=q_i\,{}_i\hat{r}_{ij}\,{}_i\hat{p}_{ii};$$

using this in (26) we obtain

(27) $$\hat{p}_{ij}=\hat{p}_{ii}\,q_i\,{}_i\hat{r}_{ij}$$

or

$$p_{ij}(t)=\int_0^t p_{ii}(s)\,q_i\,{}_ir_{ij}(t-s)\,ds.$$

Comparing this with (12.15) we see that for a stable i and $i\neq j$:

(28) $$q_i\,{}_ir_{ij}=g_{ij}.$$

Notes. For general results of the type of Theorem 1 see Chung and Doob [1]. Theorem 2 is given in Chung [5] but the proof has been strengthened. Equation (6) is a case of the "shift" transformation (cf. § I.17) which is not explicitly introduced here; to do so it would have been necessary to make a certain convention regarding the probability space (see e.g. Dynkin [2]). Corollary 2 to Theorem 2 was stated by Lévy [1] as obvious and much used, including his proof of Theorem 3; for the latter cf. Chacon [1]. Theorems 4 to 8 are from Chung [5]. For results concerning the higher derivatives of p_{ij} see Austin [1] and Jurkat [2]. The last part of this section is new.

§ 16. Imbedded renewal process

In the preceeding section we were concerned with the transition *from* a stable state i; in this section we are concerned with the transition *to* a stable state j. From the analytical point of view, the contrast is that between Theorems 3.1 and 3.2. However, there will be a difference in the presentation. In § 15 it we have developed the main results probabilistically without the aid of Theorem 3.1; here recourse will be made to the corresponding Theorem 3.2 or Theorem 12.5.

Let j be a stable but not absorbing state and consider the sequence of random variables introduced in § 15, the successive entrance times into the state j:

$$(1) \qquad \tau_1(j),\ \tau_2(j),\ \ldots,\ \tau_n(j),\ \ldots.$$

We do not assume that j is recurrent so that the domain of finiteness of $\tau_n(j)$ is E_n such that $\mathsf{E}_n \supset \mathsf{E}_{n+1}$, $n \geqq 1$. On E_n the distribution of $\tau_{n+1}(j)$ is given by

$$F_{jj}(t) = \boldsymbol{P}\{\tau_{n+1}(j;\omega) - \tau_n(j,\omega) \leqq t \mid \mathsf{E}_n\};$$

and the distribution of $\tau_1(j)$ is given by $\sum_i p_i F_{ij}(t)$. The random variables $\tau_n - \tau_{n-1}$, $n \geqq 1$, where $\tau_0 \equiv 0$, are *independent whenever finite* in the sense that for every t_n, $1 \leqq n \leqq N$, we have

$$\begin{aligned} &\boldsymbol{P}\{\tau_n(j;\omega) - \tau_{n-1}(j;\omega) \leqq t_n,\ 1 \leqq n \leqq N\} \\ &\qquad = \boldsymbol{P}\{\tau_1(j;\omega) \leqq t_1\} \prod_{n=2}^{N} \boldsymbol{P}\{\tau_n(j;\omega) - \tau_{n-1}(j;\omega) \leqq t_{n-1} \mid \mathsf{E}_n\} \\ &\qquad = \Big(\sum_i p_i F_{ij}(t_1)\Big) \prod_{n=2}^{N} F_{jj}(t_n). \end{aligned}$$

Writing $\Delta_i = \{\omega\colon x(0,\omega) = i\}$ as before the conditional distribution of $\tau_n(j)$ under the hypothesis Δ_i is therefore

$$[\delta_{ij}\boldsymbol{\epsilon} + (1 - \delta_{ij}) F_{ij}] * F_{jj}^{(n-1)}$$

where $*$ denotes the usual convolution; $\boldsymbol{\epsilon}(t)$ is *the unit distribution*, i.e. $\boldsymbol{\epsilon}(t) = 0$ or 1 according as $t < 0$ or $t \geqq 0$; $F_{jj}^{0*} = \boldsymbol{\epsilon}$ and $F_{jj}^{n*} = F_{jj}^{(n-1)*} * F_{jj}$ is the n-th iterated convolution of F_{jj} with itself. It follows that

$$(2) \qquad \sum_{n=1}^{\infty} \boldsymbol{P}\{\tau_n(j;\omega) \leqq t \mid \Delta_i\} = U_{ij}(t) = \delta_{ij}\,\boldsymbol{\epsilon}(t) + \sum_{n=0}^{\infty} [F_{ij} * F_{jj}^{n*}](t)$$

is finite for every finite $t \geqq 0$. We have for an arbitrary i,

$$(3) \qquad \left\{ \begin{aligned} \boldsymbol{P}\{x(t,\omega) = j \mid \Delta_i\} &= \sum_{n=1}^{\infty} \boldsymbol{P}\{\tau_n(j;\omega) \leqq t < \tau_{n+1}(j;\omega);\ x(t,\omega) = j \mid \Delta_i\} \\ &= \sum_{n=1}^{\infty} \boldsymbol{P}\{\tau_n(j;\omega) \leqq t;\ \lambda_n(j;\omega) > t - \tau_n(j;\omega) \mid \Delta_i\}. \end{aligned} \right.$$

Hence we obtain by Theorem 15.2,

$$(4) \qquad p_{ij}(t) = \int_0^t e^{-q_j(t-s)}\, d\,U_{ij}(s).$$

The sequence (1) corresponds to a *renewal process* (see Doob [3]). Let

$$(5) \qquad V_{ij}(t) = U_{ij}(t) - \delta_{ij}\,\boldsymbol{\epsilon}(t) = \sum_{n=0}^{\infty} [F_{ij} * F_{jj}^{n*}](t).$$

Then $V_{ij}(t)$ is the expected number of renewals of j, namely the expected number of positive $\tau_n(j)$ not exceeding t. By Theorem 12.5, F_{ij} and F_{jj} have continuous derivatives. It follows that the same is true of V_{ij}:

$$V_{ij}(t)=\int_0^t v_{ij}(s)\,ds \tag{6}$$

where v_{ij} is continuous. This result can also be deduced from Theorem 3.2 as follows. Since F_{jj} has a continuous derivative by Theorem 15.7, V_{ij} is certainly absolutely continuous so that (6) holds for some integrable v_{ij}. It then follows from (4), (5) and (6) that

$$p_{ij}(t)=\delta_{ij}\,e^{-q_j t}+\int_0^t e^{-q_j(t-s)}\,v_{ij}(s)\,ds. \tag{7}$$

Now Theorem 3.2 asserts that p_{ij} has a continuous derivative and consequently we see from (7) that v_{ij} is equal a.e. to a continuous function. We may substitute the latter into (6) and conclude that V_{ij} has indeed an everywhere continuous derivative. From now on $v_{ij}=V'_{ij}$ will denote this continuous function; for $i\neq j$ it is the *renewal density* of the state j, under the hypothesis Δ_i. It follows from (7) that

$$p'_{ij}(t)=v_{ij}(t)-p_{ij}(t)\,q_j. \tag{8}$$

This is the counterpart of (15.11). The equation

$$v_{ij}(t+s)=\sum_k p_{ik}(t)\,v_{kj}(s), \tag{9}$$

valid for $t\geqq 0$, $s>0$, is the counterpart of (15.13). It is established in the proof of Theorem 3.2 and equivalent to (3.11). We summarize these results, partly already contained in the proof of Theorem 3.2, as follows.

Theorem 1. Let j be stable and i arbitrary. The function V_{ij} defined in (5) has a continuous derivative v_{ij} in $[0, \infty)$ satisfying the equations (7), (8) and (9). Furthermore F_{ij} has a continuous derivative in $[0, \infty)$.

The last assertion follows from (13.3). Our next theorem is a form of the *renewal theorem*; it is a special case of the Corollary to Theorem 12.8.

Theorem 2. If $q_j<\infty$, then

$$\lim_{t\to\infty} v_{ij}(t)=\pi_{ij}\,q_j$$

where $\pi_{ij}=\lim\limits_{t\to\infty} p_{ij}(t)$. Equivalently, $\lim\limits_{t\to\infty} p'_{ij}(t)=0$.

Proof. Writing (4) as

$$p_{ij}(t)\,e^{q_j t}=\int_0^t e^{q_j s}\,dU_{ij}(s)$$

we see that if $t>0$, $h>0$,

$$e^{q_j t}[U_{ij}(t+h)-U_{ij}(t)] \leqq p_{ij}(t+h)\, e^{q_j(t+h)} - p_{ij}(t)\, e^{q_j t}$$

and consequently

$$U_{ij}(t+h)-U_{ij}(t) \leqq e^{q_j h}. \tag{10}$$

From this it follows that $\overline{\lim\limits_{t\to\infty}}\, v_{ij}(t)<\infty$. For otherwise let $t_n\to\infty$ and $\lim\limits_{t\to\infty} v_{ij}(t_n)=\infty$. We have from (9)

$$v_{ij}(s+t_n) \geqq p_{ii}(s)\, v_{ij}(t_n)$$

and so

$$U_{ij}(t_n+h)-U_{ij}(t_n)=\int_0^h v_{ij}(s+t_n)\, ds \geqq \int_0^h p_{ii}(s)\, ds \cdot v_{ij}(t_n).$$

Letting $n\to\infty$ we have a contradiction with (10).

We now show that $\lim\limits_{t\to\infty} v_{ij}(t)$ exists. Let $t_n\to\infty$, $t_n'\to\infty$, $a_{ij}=\lim\limits_{n\to\infty} v_{ij}(t_n)$, $a_{ij}'=\lim\limits_{n\to\infty} v_{ij}(t_n')$. Without loss of generality we may suppose that $0<t_n-t_n'\to\infty$; we have then

$$v_{ij}(t_n)=\sum_k p_{ik}(t_n-t_n')\, v_{kj}(t_n').$$

Letting $n\to\infty$ und using FATOU's lemma we obtain

$$a_{ij} \geqq \sum_k \pi_{ik}\, a_{kj}'. \tag{11}$$

The sequence $\{t_n\}$ being arbitrary it follows in particular that

$$a_{ij}' \geqq \sum_k \pi_{ik}\, a_{kj}'. \tag{12}$$

Multiplying by π_{ji} and summing over i, we have by Theorem 10.2,

$$\sum_i \pi_{ji}\, a_{ij}' \geqq \sum_i \pi_{ji} \sum_k \pi_{ik}\, a_{kj}' = \sum_k \pi_{jk}\, a_{kj}'.$$

There is thus equality here and so also in (12). Comparing with (11) we have $a_{ij}\geqq a_{ij}'$, and therefore by symmetry $a_{ij}=a_{ij}'$. Thus $\lim\limits_{t\to\infty} v_{ij}(t)$ exists and is finite. It now follows from (8) and Theorem 10.1 that $\lim\limits_{t\to\infty} p_{ij}'(t)$ exists; this limit must be zero since $p_{ij}(t)\leqq 1$. ☐

The idea behind the basic formula (3) is of course that "$x(t,\omega)=j$" amounts to "t belonging to a j-interval of $x(\cdot,\omega)$". One may inquire further about the probability that this j-interval be preceded by a k-interval where k is a stable state; in other words, the probability that the last discontinuity of the sample function before t be a jump from k to j. To find this we first note the following analogue of (2):

$$\sum_{n=1}^{\infty} \boldsymbol{P}\{\tau_n'(j;\omega)\leqq t \mid \Delta_i\} = [U_{ij} * e_{q_j}](t) = W_{ij}(t) \tag{13}$$

where from now on we use the notation

$$e_q(t) = 1 - e^{-qt}, \qquad t \geqq 0. \tag{14}$$

Formula (13) follows from (2) and Theorem 15.2. Upon differentiation we obtain

$$W'_{ij}(t) = q_j \int_0^t e^{-q_j(t-s)}\, d\,U_{ij}(s) = p_{ij}(t)\, q_j. \tag{15}$$

The desired probability can now be evaluated as follows, using Theorem 9.5 and (11.17) by taking the α there to be $\tau'_n(k)$ restricted to the set $\{\omega\colon x(\tau'_n(\omega), \omega) = j\}$, whose probability is given by Theorem 15.6:

$$\begin{aligned}
&\sum_{n=1}^{\infty} \boldsymbol{P}\{\tau'_n(k;\omega) \leqq t < \tau'_{n+1}(k;\omega);\, x(u,\omega) \equiv j,\, \tau'_n(k;\omega) \leqq u \leqq t \,|\, \Delta_i\} \\
&\quad = \sum_{n=1}^{\infty} \int_0^t \boldsymbol{P}\{x(u,\omega) \equiv j,\, \tau'_n(k;\omega) \leqq u \leqq t \,|\, \tau'_n(k) = s\}\, d_s\, \boldsymbol{P}\{\tau'_n(k;\omega) \leqq s\} \\
&\quad = \int_0^t \frac{q_{kj}}{q_k}\, e^{-q_j(t-s)}\, d\,W_{ik}(s) = \int_0^t p_{ik}(s)\, e^{-q_j(t-s)}\, ds.
\end{aligned}$$

Heavy machinery has been used in the above calculation which is given here only as an illustration of the method. It turns out that there is a more delicate proof which extends the result to an instantaneous k. Let us say that the sample function $x(\cdot,\omega)$ has a *pseudo-jump* at τ *from* k *to* j iff

$$\varliminf_{s \uparrow \tau} x(s,\omega) = k, \qquad \varliminf_{s \downarrow \tau} x(s,\omega) = j \tag{16}$$

where $k \neq j$ and either k or j may be ∞. If either k or j is stable then the corresponding $\varliminf$ may be replaced by lim by Theorem 7.4 (a), and if both are stable then the pseudo-jump becomes an ordinary jump.

Theorem 3. Let i, k be arbitrary, j stable and $\boldsymbol{P}(\Delta_i) = 1$. The probability that $x(t,\omega) = j$ and that the last discontinuity before t of the sample function $x(\cdot,\omega)$ is a pseudo-jump from k to j is equal to

$$\int_0^t p_{ik}(s)\, q_{kj}\, e^{-q_j(t-s)}\, ds \tag{17}$$

if $k \neq \infty$; and equal to

$$p_{ij}(t) - \delta_{ij}\, e^{-q_j t} - \sum_{k \neq j} \int_0^t p_{ik}(s)\, q_{kj}\, e^{-q_j(t-s)}\, ds \tag{18}$$

if $k = \infty$.

Proof. Let Δ denote the set of ω such that t belongs to a j-interval of $x(\cdot,\omega)$ and that at the left endpoint $\tau = \tau(\omega)$ of this interval (16) is true. It follows from Theorems 5.7 and 7.4 that $\boldsymbol{P}(\Delta)$ is equal to the required probability. Suppose first that $k \neq \infty$. Then the τ just defined satisfies the conditions of Theorem 12.1, indeed only (ii) needs a simple verification. Define τ_m as in that theorem and let Λ_m and Λ'_m be respectively the intersection of Λ with $\{\omega : x(\tau_m(\omega), \omega) = k\}$ and $\{\omega : x(\tau_m(\omega), \omega) \neq k\}$. We have $\lim\limits_{m \to \infty} P(\Lambda'_m)$ by Theorem 12.1. Furthermore, suppose that for infinitely many values of m tending to infinity there is an $n_m(\omega)$ such that $x(n_m 2^{-m}, \omega) = k$ and $x(s, \omega) \equiv j$ for $(n_m + 1) 2^{-m} \leqq s \leqq t$, then the corresponding set of $\{n_m 2^{-m}\}$ has a right limiting point $\tau(\omega)$ for which (16) is true by well-separability and Theorem 6.1, and consequently $\omega \in \Lambda$ except for a null set. This means, if we put

$$\mathsf{M}_m = \bigcup_{(n+1)\,2^{-m} < t} \{\omega : x(n 2^{-m}, \omega) = k;\ x(s, \omega) \equiv j,\ (n+1) 2^{-m} \leqq s \leqq t\},$$

we have $\limsup\limits_{m} \mathsf{M}_m \subset \Lambda$ so that

$$\overline{\lim_{m \to \infty}}\, \boldsymbol{P}(\mathsf{M}_m) \leqq \boldsymbol{P}(\Lambda).$$

On the other hand it is clear that

$$\Lambda_m \cap \{\omega : \tau(\omega) < t - 2^{-m}\} \subset \mathsf{M}_m.$$

Thus

$$\Lambda \cap \{\omega : \tau(\omega) < t - 2^{-m}\} \subset \mathsf{M}_m \cup \Lambda'_m$$

and consequently

$$\boldsymbol{P}(\Lambda) \leqq \underline{\lim}_{m \to \infty}\, \boldsymbol{P}(\mathsf{M}_m) + \underline{\lim}_{m \to \infty}\, \boldsymbol{P}(\Lambda'_m) = \underline{\lim}_{m \to \infty}\, \boldsymbol{P}(\mathsf{M}_m).$$

We here therefore proved that

$$\boldsymbol{P}(\Lambda) = \lim_{m \to \infty} \boldsymbol{P}(\mathsf{M}_m).$$

Since

$$\boldsymbol{P}(\mathsf{M}_m) = \sum_{(n+1)\,2^{-m} < t} p_{ik}(n 2^{-m})\, p_{kj}(2^{-m})\, e^{-q_j[t - (n+1)\,2^{-m}]}$$

and $\lim\limits_{m \to \infty} p_{kj}(2^{-m})\, 2^m = q_{kj} < \infty$ by Theorem 2.5, the sum above converges to (17) by the theory of Riemann integration. This proves the theorem in the case $k \neq \infty$. The case $k = \infty$ then follows from Theorem 7.4 by an enumeration of the possibilities, if we note also that the probability that $x(t, \omega) = j$ and that there is no discontinuity of $x(\cdot, \omega)$ before t is $\delta_{ij}\, e^{-q_j t}$. ▯

We give the counterpart of Theorem 3 here for the sake of comparison, although it belongs properly to § 15.

Theorem 4. Let j, k be arbitrary, i stable and $\boldsymbol{P}(\Delta_i) = 1$. The probability that $x(t, \omega) = j$ and that the first discontinuity (after 0) of the sample function $x(\cdot, \omega)$ is a pseudo-jump from i to k is equal to

$$(19)\qquad \int_0^t e^{-q_i(t-s)} q_{ik}\, p_{kj}(s)\, ds$$

if $k \neq \infty$; and equal to

$$(20)\qquad p_{ij}(t) - \delta_{ij}\, e^{-q_i t} - \sum_{k \neq i} \int_0^t e^{-q_i(t-s)} q_{ik}\, p_{kj}(s)\, ds$$

if $k = \infty$.

This may be proved either as a consequence of Theorem 15.2 and 15.6, or directly as in the preceding proof but more easily. The significance of both Theorems 3 and 4 will be analysed in the next section.

An instructive connection between the quantities introduced in §§ 15 and 16 is as follows. By the definition of W_{ii} in (13) and that of ${}_i r_{ij}$ in (15.19), we have on account of Theorem 15.2, if i is stable and $i \neq j$:

$$p_{ij}(t) = \int_0^t {}_i r_{ij}(t-s)\, d\, W_{ii}(s) = \int_0^t p_{ii}(s)\, q_i\, {}_i r_{ij}(t-s)\, ds,$$

the second equation by (15). This is (15.27).

Notes. The developments leading to Theorem 1 were given in Chung [5] under the hypothesis that all $q_k < \infty$. The relaxation of this to only $q_j < \infty$ is immediate upon the subsequent proof of Theorem 3.2 Theorem 2 is new. Theorems 3 and 4 under more stringent conditions are due to Doob [2], when all states are stable. The general case of Theorem 4 is considerably harder since it really involves reversing the process as in § 12.

§ 17. The two systems of differential equations

Let us recall the dual equations: if $q_i < \infty$,

$$(3.5\text{ bis})\qquad p'_{ij}(s+t) = \sum_k p'_{ik}(s)\, p_{kj}(t);$$

and if $q_j < \infty$.

$$(3.11\text{ bis})\qquad p'_{ij}(t+s) = \sum_k p_{ik}(t)\, p'_{kj}(s);$$

both valid for $s > 0$, $t \geqq 0$. These equations are valid for arbitrary i and j by Theorem 12.8 but we shall be concerned only with the cases specifically noted. The limiting cases for $s = 0$ may be written as

$$(1_{ij})\qquad p'_{ij}(t) = -q_i\, p_{ij}(t) + \sum_{k \neq i} q_{ik}\, p_{kj}(t);$$

$$(2_{ij})\qquad p'_{ij}(t) = -p_{ij}(t)\, q_j + \sum_{k \neq j} p_{ik}(t)\, q_{kj}.$$

These equations are no longer necessarily valid, but the corresponding inequalities with "$\geqq$" replacing the "$=$", which will be referred to as (1^*_{ij}) and (2^*_{ij}) respectively, are valid. The simplest way to prove these inequalities is to write for $t\geqq 0$, $h>0$,

$$\frac{p_{ij}(t+h)-p_{ij}(t)}{h}=\frac{p_{ii}(h)-1}{h}\,p_{ij}(t)+\sum_{k\neq i}\frac{p_{ik}(h)}{h}\,p_{kj}(t);$$

$$\frac{p_{ij}(t+h)-p_{ij}(t)}{h}=p_{ij}(t)\,\frac{p_{jj}(h)-1}{h}+\sum_{k\neq j}p_{ik}(t)\,\frac{p_{kj}(h)}{h};$$

let $h\downarrow 0$ and use Theorems 2.4 and 2.5 together with FATOU's lemma. We shall call the equations (1_{ij}) for a fixed stable i and all j the *first subsystem* (1_i); the equations (2_{ij}) for a fixed stable j and all i the second *subsystem* (2_j); the collection of (1_i) for all i the *first system*; and the collection of (2_j) for all j the *second system*. As a mnemonic aid we note that in the first [second] system all the q's appearing in one equation have the same first [second] index.

The more meaningful probabilistic counterparts of (3.5) and (3.11) are given by (15.13) and (16.9), which we record as

(15.13 bis) $\quad r_{ij}(s+t)=\sum_k r_{ik}(s)\,p_{kj}(t), \quad$ if $0<q_i<\infty$;

(16.9 bis) $\quad v_{ij}(t+s)=\sum_k p_{ik}(t)\,v_{kj}(s), \quad$ if $0\leqq q_j<\infty$;

valid for $s>0$, $t\geqq 0$. These equations have the analytical advantage of involving only nonnegative quantities. Applying FATOU's lemma we obtain the limiting inequalities for $s=0$:

(3^*_{ij}) $\quad r_{ij}(t)\geqq\sum_k r_{ik}(0)\,p_{kj}(t);$

(4^*_{ij}) $\quad v_{ij}(t)\geqq\sum_k p_{ik}(t)\,v_{kj}(0).$

The corresponding equations will be referred to as (3_{ij}) and (4_{ij}) which are equivalent to (1_{ij}) and (2_{ij}) respectively; smilarly for the systems (3_i) and (4_j) The equivalence is established by (15.11) and (16.8) which we record as

(15.11 bis) $\quad r_{ij}(t)=q_i^{-1}p'_{ij}(t)+p_{ij}(t), \quad$ if $0<q_i<\infty$;

(16.8 bis) $\quad v_{ij}(t)=p'_{ij}(t)+p_{ij}(t)\,q_j, \quad$ if $0\leqq q_j<\infty$;

both valid for $t\geqq 0$.

The following theorem shows that the equations of each subsystem are strongly linked in time and also to one another.

Theorem 1. If the equation (1_{ij_0}) [or $(2_{i_0 j})$] holds for a particular $t=t_0>0$, then (1_{ij}) [or (2_{ij})] holds for all j [or i] such that $j\rightsquigarrow j_0$ [or $i_0\rightsquigarrow i$] and all t.

Proof. The situation regarding the two subsystems being symmetric we will give the proof for the second subsystem or its equivalent (4_j). Suppose then that

$$v_{i_0 j}(t_0) = \sum_k p_{i_0 k}(t_0)\, v_{kj}(0). \tag{5}$$

We have by (16.9 bis) and (4_{ij}^*), if $0 \leqq t < t_0$,

$$\begin{aligned} v_{i_0 j}(t_0) &= \sum_k \Big[\sum_i p_{i_0 i}(t_0 - t)\, p_{ik}(t)\Big] v_{kj}(0) \\ &= \sum_i p_{i_0 i}(t_0 - t) \sum_k p_{ik}(t)\, v_{kj}(0) \leqq \sum_i p_{i_0 i}(t_0 - t)\, v_{ij}(t) = v_{i_0 j}(t_0). \end{aligned}$$

Hence the inequality above must be an equality. Now let $i_0 \rightsquigarrow i$; then $p_{i_0 i}(t_0 - t) > 0$ by Theorem 1.6 and the terms with subscript i in the third equation above must be equal. Hence (4_{ij}) holds for $0 \leqq t < t_0$, and consequently (4_{kj}) holds in the same range for all k such that $i \rightsquigarrow k$. Using this, we have next for such a t and any $t_1 \geqq 0$,

$$\begin{aligned} v_{ij}(t + t_1) &= \sum_k p_{ik}(t_1) \sum_l p_{kl}(t)\, v_{lj}(0) \\ &= \sum_l \Big[\sum_k p_{ik}(t_1)\, p_{kl}(t)\Big] v_{lj}(0) = \sum_l p_{il}(t + t_1)\, v_{lj}(0). \end{aligned}$$

Hence (4_j) holds for all t, proving the theorem. ▯

Corollary. If the first [or second] system (1_i) [or (2_j)] holds for a particular $t > 0$ it folds for all t.

For the first subsystem we have the following simple analytical criterion.

Theorem 2. If the first subsystem (1_i) holds for a particular $t > 0$, then we have

$$\sum_{k \neq i} q_{ik} = q_i. \tag{6}$$

Conversely, if (6) is true then (1_i) holds for all t.

Proof. The case $q_i = 0$ being trivial we may suppose that $0 < q_i < \infty$ and consider the equivalent subsystem (3_i). Suppose that the latter holds for a certain $t > 0$. Summing over j we obtain by (15.12)

$$1 = \sum_j r_{ij}(t) = \sum_j \sum_k r_{ik}(0)\, p_{kj}(t) = \sum_k r_{ik}(0). \tag{7}$$

This is equivalent to (6) by (15.14). Conversely, we observe that if the middle "$=$" above is replaced by "$\geqq$" the resulting inequality holds always by (3_i^*). Now if (6) is true then the extreme terms of (7) are equal and so for each $t > 0$ there must be equality in (3_{ij}^*) for all j, proving (3_i). ▯

For the second subsystem there does not appear to be an analogous criterion; instead we prove the following result.

Theorem 3. Let

$$d_{ij}(t) = v_{ij}(t) - \sum_k p_{ik}(t)\, v_{kj}(0) = p'_{ij}(t) - \sum_k p_{ik}(t)\, q_{kj}.$$

Then we have if $t \geqq 0$, $s > 0$,

$$d_{ij}(t+s) = \sum_k p_{ik}(t)\, d_{kj}(s). \tag{8}$$

The function d_{ij} is nonnegative and continuous in $(0, \infty)$; $\lim_{t \downarrow 0} d_{ij}(t) = 0$; $\lim_{t \to \infty} d_{ij}(t)$ exists and is finite.

Proof. The equation (8) follows easily from (16.9 bis); $d_{ij} \geqq 0$ by (4^*_{ij}); d_{ij} is continuous in $(0, \infty)$ by Theorem 2.3; the limit at zero exists by the same theorem and the limit at infinity exists and is finite as in the proof of Theorem 16.2. Moreover we have

$$\overline{\lim_{t \downarrow 0}}\, v_{ij}(t) \geqq \lim_{t \downarrow 0} \sum_k p_{ik}(t)\, v_{kj}(0) + \overline{\lim_{t \downarrow 0}}\, d_{ij}(t).$$

Hence it follows from Fatou's lemma that

$$(1 - \delta_{ij})\, q_{ij} \geqq (1 - \delta_{ij})\, q_{ij} + \overline{\lim_{t \downarrow 0}}\, d_{ij}(t),$$

hence $\lim_{t \downarrow 0} d_{ij}(t) = 0$. ▯

Corollary. If $q_j < \infty$, then $\sum_k p_{ik}(t)\, q_{kj}$ is continuous in $(0, \infty)$ and tends to q_{ij} as $t \downarrow 0$. The result remains true if (p_{ij}) is a standard substochastic transition matrix.

The last remark follows from Theorem 3.3 and will be used later.

We can now give a necessary and sufficient condition for the validity (for all $t > 0$) of *each equation* (1_{ij}) or (2_{ij}) in terms of sample function discontinuities.

Theorem 4. The equation (1_{ij}) is valid if and only if for every $t > 0$, under the hypothesis that $x(0, \omega) = i$ and $x(t, \omega) = j$ the probability is one that the first discontinuity of the sample function, where existing, is not a pseudo-jump to ∞. The equation (2_{ij}) is valid if and only if for every $t > 0$, under the hypothesis that $x(0, \omega) = i$ and $x(t, \omega) = j$ the probability is one that the last discontinuity of the sample function before t, where existing, is not a pseudo-jump from ∞. The subsystem (1_i) is valid if and only if under the hypothesis $x(0, \omega) = i$ the probability is one that the first discontinuity of the sample function, where existing, is not a pseudo-jump to ∞. The subsystem (2_j) is valid if and only if for every $t > 0$, under the hypothesis $x(t, \omega) = j$ the probability is one that the last discontinuity of the sample function before t, where existing, is not a pseudo-jump from ∞.

Proof. The stated condition for (1_{ij}) is equivalent by Theorem 16.4 to the validity of the following equation:

$$(9)\qquad p_{ij}(t)=\delta_{ij}\,e^{-q_i t}+\sum_{k\neq i}\int_0^t e^{-q_i(t-s)}\,q_{ik}\,p_{kj}(s)\,ds.$$

The series converges uniformly since $\sum_{k\neq i} q_{ik}\leqq q_i$; hence multiplication of (9) by $e^{q_i t}$ shows its equivalence to (1_{ij}), proving the first assertion. Similarly, the stated condition for (2_{ij}) is equivalent by Theorem 16.3 to the validity of the following equation:

$$(10)\qquad p_{ij}(t)=\delta_{ij}\,e^{-q_j t}+\sum_{k\neq j}\int_0^t p_{ik}(s)\,q_{kj}\,e^{-q_j(t-s)}\,ds.$$

Here the equivalence of (10) and its differentiated form (2_{ij}) is justified by the well known theorem on termwise integration of nonnegative functions together with the continuity of the differentiated series given by the Corollary to Theorem 3 above. This proves the second assertion. The rest of the theorem follows immediately. ▯

Theorem 5. Suppose that all $q_k<\infty$. A necessary and sufficient condition that both the first and second system be valid is that almost all sample functions have the following property: if $x(s,\omega)\to\infty$ as $s\to t$ from one side then the same is true as $s\to t$ from both sides.

Proof. Since all states are stable, each fixed t belongs to a stable interval with probability one by Theorem 5.7. Suppose all sample functions have the stated property. Then for each t, the first discontinuity after t and the last discontinuity before t of almost every sample function must be a jump to and from a state $\neq\infty$. Hence both systems are valid by Theorem 4. Conversely, suppose that there exists a set Λ_1 with $\boldsymbol{P}(\Lambda_1)>0$ such that if $\omega\in\Lambda_1$, the sample function $x(\cdot,\omega)$ does not have the stated property. For almost every such ω there exists a stable interval of $x(\cdot,\omega)$ such that the first discontinuity following the interval or the last discontinuity preceding the interval is a pseudo-jump to or from ∞. By separability there exists a fixed r in the separability set, a fixed $k\neq\infty$, and a set $\Lambda_2\subset\Lambda_1$ with $\boldsymbol{P}(\Lambda_2)>0$ such that if $\omega\in\Lambda_2$, then $x(r,\omega)=k$ and the first discontinuity after r or the last discontinuity before r of $x(\cdot,\omega)$ is a pseudo-jump to or from ∞. Hence either (1_k) or (2_k) must fail by Theorem 4. ▯

Suppose that $0<q_i<\infty$ for all i. Let us now define new probabilites as follows:

$$(11)\qquad \begin{cases} p_{ij}^{\langle n\rangle}(t)=\boldsymbol{P}\{x(\cdot,\omega)\text{ has exactly } n \text{ discontinuities}\\ \qquad\qquad\text{in } (0,t);\ x(t,\omega)=j\,|\,x(0,\omega)=i\}.\end{cases}$$

Note that it is immaterial how we count the discontinuity at 0 or t, since the probability of a discontinuity at a fixed point is zero when all states are stable, by Theorem 6.2. We have clearly

$$p_{ij}^{\langle 0\rangle}(t) = \delta_{ij}\, e^{-q_i t} = \delta_{ij}\, e^{-q_j t}; \tag{12}$$

and using the arguments for Theorems 16.3 and 16.4 we obtain the dual recursive formulas for $n \geqq 0$:

$$p_{ij}^{\langle n-1\rangle}(t) = \sum_{k \neq i} \int_0^t e^{-q_i(t-s)} q_{ik}\, p_{kj}^{\langle n\rangle}(s)\, ds; \tag{13}$$

$$p_{ij}^{\langle n+1\rangle}(t) = \sum_{k \neq j} \int_0^t p_{ik}^{\langle n\rangle}(s)\, q_{kj}\, e^{-q_j(t-s)}\, ds. \tag{14}$$

Put also

$$\overline{p}_{ij}(t) = \sum_{n=0}^{\infty} p_{ij}^{\langle n\rangle}(t) \tag{15}$$

so that

$$\begin{cases} \overline{p}_{ij}(t) = \boldsymbol{P}\{x(\cdot,\omega) \text{ has only a finite number of discontinuities} \\ \qquad\qquad \text{in } (0,t);\ x(t,\omega) = j \mid x(0,\omega) = i\}. \end{cases} \tag{16}$$

We have then, dually,

$$\overline{p}_{ij}(t) = \delta_{ij}\, e^{-q_i t} + \sum_{k \neq i} \int_0^t e^{-q_i(t-s)}\, q_{ik}\, \overline{p}_{kj}(s)\, ds; \tag{17}$$

$$\overline{p}_{ij}(t) = \delta_{ij}\, e^{-q_j t} + \sum_{k \neq j} \int_0^t \overline{p}_{ik}(s)\, q_{kj}\, e^{-q_j(t-s)}\, ds. \tag{18}$$

The last two formulas can be established directly by the arguments for Theorems 16.3 and 16.4. For a later purpose we observe however that they also follow analytically from (12)—(15). It is easily seen from the definition that $(\overline{p}_{ij})$, $i, j \in \boldsymbol{I}$, is a substochastic transition matrix (see §3). Furthermore all the equations (1_{ij}) and (2_{ij}) hold if (p_{ij}) is replaced by $(\overline{p}_{ij})$ (see Corollary to Theorem 3 for the differentiation of (18)); in other words $(\overline{p}_{ij})$ gives a solution of both system of differential equations. It thus appears that for a study of these equations we should consider solutions which form a substochastic transition matrix. We shall proceed to do this in the next section.

Notes. The "first" and "second" systems are usually called the "backward" and "forward" systems respectively. The latter modifiers refer to differentiation with respect to the "earlier" or "later" time in the nonstationary transition case. But here they are somewhat confusing in view of Theorems 16.3 and 16.4, since the nearest discontinuity forward [backward] yields the backward [forward] equation. The Corollary to Theorem 1 is contained in CHUNG [5] but the theorem was orally communicated by AUSTIN. Theorems 2 and 4 are somewhat strengthened forms of results due to DOOB [2]. Theorems 3 and 5 are new. An excellent introduction to the semigroup theory of the two systems of differential equations is given in REUTER [2].

§ 18. The minimal solution

In the sequel we shall use the alternative notation

$$q_{ii} = -q_i.$$

The point of view in § 17 regarding the two systems of differential equations is as follows. The standard transition matrix (p_{ij}) being given, the matrix $(q_{ij}) = (p'_{ij}(0))$ is thereby determined. Assuming that $q_{ii} > -\infty$ for every i we consider the validity of the equations (1_{ij}) and (2_{ij}) of § 17. The elements of the matrix (q_{ij}) satisfy the following conditions:

$$(1) \qquad q_{ij} \geqq 0 \quad \text{if } i \neq j; \qquad -\infty < \sum_j q_{ij} \leqq 0 \quad \text{for every } i.$$

A matrix of constants (q_{ij}) satisfying (1) will be called a *Q-matrix*; it is called *conservative* iff the last "$\leqq$" is replaced by "$=$", namely iff (17.6) holds for every i. The matrix $(p'_{ij}(0))$ will be called *the Q-matrix* of a substochastic transition matrix (p_{ij}).

In this section we take a different point of view. Given an arbitrary Q-matrix, we consider the differential equations

$$\left.\begin{aligned} (2) \qquad & z'_{ij}(t) = \sum_k q_{ik} z_{kj}(t) \\ (3) \qquad & z'_{ij}(t) = \sum_k z_{ik}(t) q_{kj} \end{aligned}\right\} \quad i, j \in \boldsymbol{I}, \; t \geqq 0$$

in the unknown functions z_{ij} and seek a solution (z_{ij}) which is a standard substochastic transition matrix. The following construction of a solution is inspired by the discussion at the end of § 17. With the given (q_{ij}) we define $p_{ij}^{\langle n\rangle}(t)$, $\bar{p}_{ij}(t)$ by (17.12), (17.13) and (17.15). Clearly $p_{ij}^{\langle n\rangle}(t) \geqq 0$. Next, if we put $\sigma_{ij}^{\langle n\rangle}(t) = \sum_{\nu=0}^{n} p_{ij}^{\langle\nu\rangle}(t)$ we have from (17.3)

$$\sigma_{ij}^{\langle n+1\rangle}(t) = \delta_{ij} e^{-q_i t} + \sum_{k \neq i} \int_0^t e^{-q_i(t-s)} q_{ik} \sigma_{kj}^{\langle n\rangle}(s)\, ds.$$

A simple induction on n then shows that

$$(4) \qquad \sum_j \sigma_{ij}^{\langle n\rangle}(t) = \sum_j \sum_{\nu=0}^{n} p_{ij}^{\langle\nu\rangle}(t) \leqq 1$$

for every t and n. Hence the series in (17.15) converges. Summing (17.13) over $n \geqq 0$ and differentiating we see that $(\bar{p}_{ij})$ is a solution of

(2). To prove that it is also a solution of (3) we shall show that if we use (17.14) instead of (17.13) in the above construction, the resulting $\breve{p}_{ij}^{\langle n\rangle}$ coincides with the previous $p_{ij}^{\langle n\rangle}$ for all n. This is true for $n=0$ by (17.12). For $n=1$ both $p_{ij}^{\langle 1\rangle}$ and $\breve{p}_{ij}^{\langle 1\rangle}$ are equal to

$$(1-\delta_{ij})\,q_{ij}\,[e_{q_i} * e_{q_j}]$$

in the notation of (16.14). Assuming that $p_{ij}^{\langle\nu\rangle}=\breve{p}_{ij}^{\langle\nu\rangle}$ for all i and j, $\nu=n-1$ and $\nu=n$, we see that both $p_{ij}^{\langle n+1\rangle}$ and $\breve{p}_{ij}^{\langle n+1\rangle}$ are equal to

$$e_{q_i} * \Big[\sum_{k\neq i}\ \sum_{l\neq j} q_{ik}\, p_{kl}^{\langle n-1\rangle}\, q_{lj}\Big] * e_{q_i}.$$

Hence the induction is complete and consequently (17.14) holds for the $p_{ij}^{\langle n\rangle}$'s constructed from (17.13). Summing (17.14) over $n\geqq 0$ and differentiating we see that (3) is satisfied with $z_{ij}=\bar{p}_{ij}$ *for almost all* t. To prove that $(\bar{p}_{ij})$ is a solution of (3) for all t we will first show that $(\bar{p}_{ij})$ is a substochastic transition matrix. For this suppose we prove the identity

$$p_{ij}^{\langle n\rangle}(s+t)=\sum_{\nu=0}^{n}\sum_{k} p_{ik}^{\langle\nu\rangle}(s)\, p_{kj}^{\langle n-\nu\rangle}(t) \tag{5}$$

in i, j, s and t. For $n=0$ this is trivial. Assuming that (5) is true for a given n, we have from (17.13):

$$\begin{aligned}
&\sum_{\nu=0}^{n+1}\sum_{k} p_{ik}^{\langle\nu\rangle}(s)\, p_{kj}^{\langle n+1-\nu\rangle}(t)\\
&\quad=\sum_{k} p_{ik}^{\langle 0\rangle}(s)\, p_{kj}^{\langle n+1\rangle}(t)+\sum_{\nu=1}^{n+1}\sum_{k}\sum_{l\neq i}\int_0^s e^{-q_i u}\, q_{il}\, p_{lk}^{\langle\nu-1\rangle}(s-u)\, p_{kj}^{\langle n+1-\nu\rangle}(t)\, du\\
&\quad=e^{-q_i s}\, p_{ij}^{\langle n+1\rangle}(t)+\sum_{l\neq i}\int_0^s e^{-q_i u}\, q_{il}\, p_{lj}^{\langle n\rangle}(s-u+t)\, du\\
&\quad=e^{-q_i s}\sum_{l\neq i}\int_0^t e^{-q_i u}\, q_{il}\, p_{lj}^{\langle n\rangle}(t-u)\, du+\sum_{l\neq i}\int_0^s e^{-q_i u}\, q_{il}\, p_{lj}^{\langle n\rangle}(s+t-u)\, du\\
&\quad=\sum_{l\neq i}\int_0^{s+t} e^{-q_i u}\, q_{il}\, p_{li}^{\langle n\rangle}(s+t-u)\, du=p_{ij}^{\langle n+1\rangle}(s+t).
\end{aligned}$$

Hence (5) is proved by induction on n. Summing it over n we obtain

$$\bar{p}_{ij}(s+t)=\sum_{k}\sum_{\nu=0}^{\infty} p_{ik}^{\langle\nu\rangle}(s)\sum_{n=\nu}^{\infty} p_{kj}^{\langle n-\nu\rangle}(t)=\sum_{k}\bar{p}_{ik}(s)\,\bar{p}_{kj}(t).$$

This establishes property (1.C) for the matrix $(\bar{p}_{ij})$; combined with (4) we see that it is indeed a substochastic transition matrix. It is standard since $\varliminf_{t\downarrow 0}\bar{p}_{ii}(t)\geqq\lim_{t\downarrow 0} p_{ii}^{\langle 0\rangle}(t)=1$. By Theorems 3.1 (or 3.2) and 3.3, $\bar{p}'_{ij}(t)$ is a continuous function of t; by the Corollary to Theorem 17.3, so is

$\sum_k \bar{p}_{ik}(t)\, q_{kj}$. We have already proved that these two functions of t are equal for a.a. t; hence they are equal everywhere. This completes the proof of the following theorem.

Theorem 1. For any Q-matrix (q_{ij}) the system of equations (2) and (3) have a solution $(\bar{p}_{ij})$ which is a standard substochastic transition matrix. The elements of the solution matrix are constructed by (17.12), either (17.13) or (17.14), and (17.15).

Let (z_{ij}) be any substochastic transition matrix such that for every i and j,

$$z'_{ij}(0) = q_{ij}; \tag{6}$$

in other words (z_{ij}) satisfies the equations (2) and (3) for $t=0$. Then z_{ij} and z'_{ij} are continuous by Theorem 3.3, and the inequalities corresponding to (1^*_{ij}) of § 17 hold for (z_{ij}) by FATOU's lemma. In particular we have $z'_{ij}(t) \geqq -q_i\, z_{ij}(t)$ from which it follows that

$$z_{ij}(t) \geqq \delta_{ij}\, e^{-q_i t} = p_{ij}^{\langle 0\rangle}(t).$$

Furthermore, by integrating the said inequalities we have (cf. (17.9))

$$z_{ij}(t) \geqq \delta_{ij}\, e^{-q_i t} + \sum_{k \neq i} \int_0^t e^{-q_i(t-s)}\, q_{ik}\, z_{kj}(s)\, ds.$$

Assuming then, for the sake of induction on n, that $z_{kj}(t) \geqq \sum_{\nu=0}^{n} p_{kj}^{\langle\nu\rangle}(t)$ for all k and j, we obtain from (17.13) and the above,

$$\begin{aligned} z_{ij}(t) &\geqq \delta_{ij}\, e^{-q_i t} + \sum_{\nu=0}^{n} \sum_{k \neq i} \int_0^t e^{-q_i(t-s)}\, q_{ik}\, p_{kj}^{\langle\nu\rangle}(s)\, ds \\ &= \delta_{ij}\, e^{-q_i t} + \sum_{\nu=0}^{n} p_{ij}^{\langle\nu+1\rangle}(t) = \sum_{\nu=0}^{n+1} p_{ij}^{\langle\nu\rangle}(t). \end{aligned}$$

This completes the induction and we have upon letting $n \to \infty$

$$z_{ij}(t) \geqq \bar{p}_{ij}(t). \tag{7}$$

Theorem 2. If (z_{ij}) is any substochastic transition matrix satisfying (6) and $(\bar{p}_{ij})$ is the solution constructed according to Theorem 1, then (7) is true.

The substochastic transition matrix $(\bar{p}_{ij})$ will be called the *minimal solution corresponding to the Q-matrix* (q_{ij}).

Theorem 3. If the minimal solution corresponding to a given Q-matrix is a transition matrix then it is the unique substochastic transition matrix whose Q-matrix coincides with the given one, necessarily conservative. In particular it is the unique solution of the two systems (2) and (3).

Proof. Let (z_{ij}) be a solution; then (7) is true by Theorem 2. It follows that

$$1 \geqq \sum_j z_{ij}(t) \geqq \sum_j \bar{p}_{ij}(t) = 1$$

and consequently $z_{ij} = \bar{p}_{ij}$ for every i and j. The Q-matrix is conservative by Theorem 17.2. ▯

The converse of this theorem will be given in the Corollary to Theorem 19.4.

If the minimal solution is substochastic, we may complete it by adding one new state ϑ in accordance with Theorem 3.3. Any M.C. with the resulting transition matrix, an arbitrary initial distribution and the minimal state space $\hat{\boldsymbol{I}}$ will be called a *minimal chain* corresponding to the given Q-matrix.

We proceed to derive some further properties of the minimal solution, leaving the probabilistic meaning to the next section. Since each $p_{ij}^{\langle n\rangle}$ is obviously continuous by induction on n, and $\bar{p}_{ij}$ is also continuous, it follows by DINI's theorem that the series in (17.15) converges uniformly in every finite interval. The convergence need not be uniform in $\boldsymbol{T}$ since if $q_j > 0$, $\lim_{t\to\infty} p_{ij}^{\langle n\rangle}(t) = 0$ for each n but $\lim_{t\to\infty} \bar{p}_{ij}(t)$ may be positive. The following result is essentially a special case of Theorem 15.3 on account of Theorem 3.3, and can also be easily proved by induction on n.

Theorem 4. Each $p_{ij}^{\langle n\rangle}$ or $\bar{p}_{ij}$ is either identically zero or never zero in $\boldsymbol{T}^0$.

For any subset J of the index set $\boldsymbol{I}$ let

$$\bar{p}_{iJ}(t) = \sum_{j\in J} \bar{p}_{ij}(t). \tag{8}$$

Theorem 5. If $\bar{p}_{i\boldsymbol{I}}(t_0) = 1$ for a certain $t_0 > 0$ then $\bar{p}_{i\boldsymbol{I}}(t) \equiv 1$ for all $t \geqq 0$; in this case $\bar{p}_{k\boldsymbol{I}}(t) \equiv 1$ for every k for which $\bar{p}_{ik}(t) \not\equiv 0$.

Proof. We have if $0 < s < t_0$,

$$\bar{p}_{i\boldsymbol{I}}(t_0) = \sum_j \bar{p}_{ij}(t_0) = \sum_k \bar{p}_{ik}(s) \sum_j \bar{p}_{kj}(t_0 - s) = \sum_k \bar{p}_{ik}(s)\, \bar{p}_{k\boldsymbol{I}}(t_0 - s). \tag{9}$$

It follows that if $\bar{p}_{i\boldsymbol{I}}(t_0) = 1$ and $\bar{p}_{ik}(s) > 0$ then $\bar{p}_{k\boldsymbol{I}}(t_0 - s) = 1$. Let E_i be the subset of $\boldsymbol{I}$ such that $k \in E_i$ if and only if $\bar{p}_{ik}(t) \not\equiv 0$; then $k \in E_i$ implies $\bar{p}_{ik}(t) > 0$ for all $t > 0$ by Theorem 4. Hence if $k \in E_i$ then $\bar{p}_{k\boldsymbol{I}}(t) = 1$ for $0 \leqq t \leqq t_0$. We have therefore if $t_0 < t \leqq 2t_0$,

$$\bar{p}_{i\boldsymbol{I}}(t) = \sum_{k\in E_i} \bar{p}_{ik}(t_0)\, \bar{p}_{k\boldsymbol{I}}(t - t_0) = \sum_{k\in E_i} \bar{p}_{ik}(t_0) = 1.$$

Repeating this argument we conclude that $\overline{p}_{i\boldsymbol{I}}(t)\equiv 1$. This proves the first assertion from which the second follows by what has already been proved. □

Corollary. The series in (8) converges uniformly in every finite interval and $\overline{p}_{iJ}$ is continuous in $\boldsymbol{T}^0$.

Proof. Since the functions $\{\overline{p}_{i\boldsymbol{I}},\, i\in\boldsymbol{I}\}$ satisfy the equations (9) which are of the form (2.4′), it follows from Theorem 3.3 that $\overline{p}_{i\boldsymbol{I}}$ is continuous in $\boldsymbol{T}_0$. Hence by Dini's theorem the series (8) with $J=\boldsymbol{I}$ converges uniformly as stated. The same is then true for an arbitrary $J\subset\boldsymbol{I}$. □

Let us put

$$L_i^{\langle n\rangle}(t)=1-\sum_{\nu=0}^{n} p_{i\boldsymbol{I}}^{\langle\nu\rangle}(t).$$

It follows from (4) that $L_i^{\langle n\rangle}(t)\geqq 0$; since each $p_{ij}^{\langle n\rangle}$ is continuous in $\boldsymbol{T}$ by the preceding Corollary, so is $L_i^{\langle n\rangle}$. Let

$$L_i(t)=\lim_{n\to\infty} L_i^{\langle n\rangle}(t)=1-\overline{p}_{i\boldsymbol{I}}(t);\tag{10}$$

then L_i is a continuous function in $\boldsymbol{T}$ by the Corollary above, but more is true as follows.

Theorem 6. Each L_i has a continuous derivative l_i satisfying for $s\geqq 0,\, t>0$:

$$l_i(s+t)=\sum_j \overline{p}_{ij}(s)\, l_j(t).\tag{11}$$

Proof. From the definition (10) and the semigroup property (1.C) of $(\overline{p}_{ij})$, we deduce for each $i\in\boldsymbol{I}$:

$$\sum \overline{p}_{ij}(s)\, L_j(t)=L_i(s+t)-L_i(s).\tag{12}$$

Applying Theorem 12.4 to the system of equations (12) we obtain (11). □

Comparison of this result with Theorem 12.5 suggests that the minimal transition probabilities may be regarded as taboo probabilities with the "ficticious taboo state ∞". This formal analogy has been substantiated and much refined by the boundary theory; see Chung [12] and [14]; Neveu [2].

Theorem 7. If the Q-matrix is conservative, then

$$L_i^{\langle n\rangle}(t)=\int_0^t \sum_j p_{ij}^{\langle n\rangle}(s)\, q_j\, ds.\tag{13}$$

Proof. Using (17.14), we have if $n \geqq 0$,

$$\begin{aligned}\int_0^t p_{ij}^{\langle n+1\rangle}(u) q_j \, du &= \int_0^t \int_0^u \sum_{k \neq j} p_{ik}^{\langle n\rangle}(s) q_{kj} \, e^{-q_j(u-s)} q_j \, ds \, du \\ &= \sum_{k \neq j} \int_0^t p_{ik}^{\langle n\rangle}(s) q_{kj} [1 - e^{-q_j(t-s)}] \, ds \\ &= \sum_{k \neq j} \int_0^t p_{ik}^{\langle n\rangle}(s) q_{kj} \, ds - p_{ij}^{\langle n+1\rangle}(t).\end{aligned}$$

Summing over j, we obtain

$$p_{i\mathbf{I}}^{\langle n+1\rangle}(t) = \int_0^t \sum_j p_{ij}^{\langle u\rangle}(s) q_j \, ds - \int_0^t \sum_j p_{ij}^{\langle n+1\rangle}(s) q_j \, ds.$$

It follows that

$$\begin{aligned}\sum_{\nu=0}^{n} p_{i\mathbf{I}}^{\langle \nu\rangle}(t) &= e^{-q_i t} + \int_0^t q_i e^{-q_i s} \, ds - \int_0^t \sum_j p_{ij}^{\langle n\rangle}(s) q_j \, ds \\ &= 1 - \int_0^t \sum_j p_{ij}^{\langle n\rangle}(s) q_j \, ds,\end{aligned}$$

proving (13). ▯

Corollary. A sufficient condition that $L_i(t) \equiv 0$ for all t is that $\sum_j \bar{p}_{ij}(t) q_j < \infty$ for some $t > 0$.

Proof. If $0 < s < t$ we have

$$\sum_j \bar{p}_{ij}(s) q_j \leqq \frac{1}{\bar{p}_{ii}(t-s)} \sum_j \bar{p}_{ij}(t) q_j$$

and consequently

$$\sum_{n=0}^{\infty} L_i^{\langle n\rangle}(t) = \int_0^t \sum_j \bar{p}_{ij}(s) q_j \, ds < \infty.$$

Hence $L_i(t) = \lim_{n \to \infty} L_i^{\langle n\rangle}(t) = 0$ and the Corollary follows from Theorem 5. ▯

Theorem 8. If the Q-matrix is conservative, then

$$q_i L_i^{\langle n+1\rangle}(t) + \sum_j p_{ij}^{\langle n+1\rangle}(t) q_j = \sum_{k \neq i} q_{ik} L_k^{\langle n\rangle}(t). \tag{14}$$

Proof. Multiplying (17.13) through by $e^{q_i t}$, differentiating, cancelling $e^{q_i t}$ and integrating, we obtain

$$p_{ij}^{\langle n+1\rangle}(t) + q_i \int_0^t p_{ij}^{\langle n+1\rangle}(s) \, ds = \sum_{k \neq i} \int_0^t q_{ik} p_{kj}^{\langle n\rangle}(s) \, ds.$$

Multiplying by q_j, summing over j and using (13), we obtain (14). ▯

An equivalent form of (14) is as follows:

$$(15)\qquad L_i^{\langle n+1\rangle}(t) = \int_0^t e^{-q_i(t-s)} \sum_{k \neq i} q_{ik} L_k^{\langle n\rangle}(s)\, ds.$$

Letting $n \to \infty$ we have

$$(16)\qquad L_i(t) = \int_0^t e^{-q_i(t-s)} \sum_{k \neq i} q_{ik} L_k(s)\, ds.$$

The differentiated form of (16) is

$$(17)\qquad l_i(t) = L_i'(t) = \sum_k q_{ik} L_k(t)$$

and the right member is equal to

$$\sum_k q_{ik} \sum_j \left(1 - \overline{p}_{kj}(t)\right) = - \sum_j \sum_k q_{ik}\, \overline{p}_{kj}(t) = - \sum_j \overline{p}_{ij}'(t).$$

Thus the series $\sum_j \overline{p}_{ij}(t) = 1 - L_i(t)$ may be differentiated term by term. This is actually a case of (3.3), in view of Theorem 3.3. Another consequence of (17) is that $l_i(0) = 0$ for each i.

It follows from (15), by induction on n, that

$$(18)\qquad \lim_{t \to \infty} L_i^{\langle n\rangle}(t) = \int_0^\infty \sum_j p_{ij}^{\langle n\rangle}(s)\, q_j\, ds = 1.$$

The probabilistic meaning of the formulas (12), (16) and (18) will become clear in the next section.

Notes. The main results of this section are specializations of those in FELLER [1] which treats more general processes. A new treatment of our special case by means of Laplace transforms (resolvents) is given in FELLER [5]; another approach through approximation by truncated matrices is given by REUTER and LEDERMANN [1]; a third one by a perturbation method by KATO [1]. See also REUTER [2] and HILLE and PHILLIPS [1]. The methods of functional analysis, however, tend to obliterate certain fine points which may be significant for the study of the stochastic process. It may be mentioned that even for the minimal solution $(\overline{p}_{ij})$, questions of representation, higher derivatives and further analytical properties seem to be open, except in special cases such as bounded or symmetric Q-matrix; see e.g. AUSTIN [1]; JURKAT [2]. For historical orientation see KOLMOGOROV [1].

§ 19. The first infinity

We return to the M. C. $\{x_t,\, t \in \boldsymbol{T}\}$ which is assumed to be well-separable and Borel measurable. Furthermore we assume that for each i, $q_i > 0$ and that $(p_{ij}'(0)) = (q_{ij})$ is a conservative Q-matrix. By Theorem 17.2

this is equivalent to the validity of the first system of differential equations. In probabilistic language, all states are stable and non-absorbing and the discontinuity of the sample functions at any exit time is an ordinary jump with probability one (see Theorems 15.2 and 15.6). It follows that we can enumerate the successive jumps until the first discontinuity that is not a jump, if such a discontinuity exists. By Theorem 7.4, since only case (c) there is possible, the sample function tends to infinity as this discontinuity is approached from the left. Let us put

$$\tau_\infty(\omega) = \inf\{t\colon t>0;\ \lim_{s\uparrow t} x(s,\omega)=\infty\} \tag{1}$$

if the t-set between the braces is not empty, and $\tau_\infty(\omega)=\infty$ otherwise; $\tau_\infty(\omega)$ is called the *first infinity* of $x(\cdot,\omega)$. It is easy to verify that τ_∞ is an optional random variable relative to $\{x_t, t\in \boldsymbol{T}\}$. For each ω, we have $\tau_\infty(\omega)=\infty$ if and only if the sample function $x(\cdot,\omega)$ is a "step function" in $\boldsymbol{T}$, namely one which has no other discontinuities except jumps on a discrete (finite or countably infinite) set of points and which is constant between jumps. Let the successive points of jump of $x(\cdot,\omega)$ until $\tau_\infty(\omega)$ be $\{\tau_n(\omega), n\geqq 1\}$; for a.a. ω this is an increasing infinite sequence and

$$\lim_{n\to\infty} \tau_n(\omega) = \tau_\infty(\omega).$$

Let $\tau_0\equiv 0$ and

$$\varrho_n(\omega) = \tau_{n+1}(\omega) - \tau_n(\omega)$$

for $n\geqq 0$ so that

$$\tau_\infty(\omega) = \sum_{n=0}^{\infty} \varrho_n(\omega). \tag{2}$$

We may now apply the analytical results of § 18 to the probabilities $p_{ij}^{\langle n\rangle}$ and $\overline{p}_{ij}$ defined in (17.11) and (17.16). The quantities $p_{i\boldsymbol{I}}^{\langle n\rangle}(t)$, $L_i^{\langle n\rangle}(t)$, $\overline{p}_{i\boldsymbol{I}}(t)$ and $L_i(t)$ are, respectively, the probabilities that $x(\cdot,\omega)$ has exactly n, more than n, finitely many, and infinitely many jumps (and no other discontinuities) in $(0,t)$ (or $[0,t]$ indifferently), under the hypothesis that $x_0(\omega)=i$. This interpretation remains valid even if (q_{ij}) is not conservative, by Theorem 15.6, since a pseudo-jump to ∞ is a discontinuity other than a jump. We have therefore

$$\begin{aligned} L_i^{\langle n\rangle}(t) &= \boldsymbol{P}\{\tau_{n+1}(\omega)\leqq t \mid x_0(\omega)=i\};\\ L_i(t) &= \boldsymbol{P}\{\tau_\infty(\omega)\leqq t \mid x_0(\omega)=i\}. \end{aligned} \tag{3}$$

According to (18.13), the distribution $L_i^{\langle n\rangle}$ of τ_{n+1} has the density function $\sum_j p_{ij}^{\langle n\rangle}(t)\,q_j$ which is continuous by (18.14). As for the distribution L_i of τ_∞, it has a continuous density function given by (18.17), which is the limit as $n\to\infty$ of the density function of τ_n, by letting $n\to\infty$ in (18.14). We have $\boldsymbol{P}\{\tau_\infty(\omega)>t|x_0(\omega)=i\}=1$ if and only if $\bar{p}_{i\boldsymbol{I}}(t)=1$ or $L_i(t)=0$. Theorem 18.5 asserts that each L_i is either identically zero or never zero. This means that if almost all sample functions with $x(0,\omega)=i$ are step functions in $(0,\varepsilon)$ for any $\varepsilon>0$ then so are they in $(0,\infty)$. Alternatively, if there is a positive probability that the first infinity is finite then there is a positive probability that it is less than any given positive number. This result is to be compared with Theorem 15.3.

For the sake of definiteness let us now take the x_+ version so that

$$x(t,\omega)=x\big(\tau_n(\omega),\omega\big) \quad \text{if} \quad \tau_n(\omega)\leqq t<\tau_{n+1}(\omega).$$

Let $\chi_n(\omega)=x\big(\tau_n(\omega),\omega\big)$, $n\geqq 0$. It follows from the strong Markov property and Theorem 15.6 that $\{\chi_n,\ n\geqq 0\}$ is a d.p.M.C. with the transition matrix (r_{ij}) defined as follows

$$r_{ij}=\begin{cases} q_{ij}/q_i & \text{if} \quad i\neq j;\\ 0 & \text{if} \quad i=j.\end{cases}$$

Note that $r_{ij}=r_{ij}(0)$; see (15.14). The d.p. one-step transition matrix (r_{ij}) will be called the *jump matrix* associated with the c.p. transition matrix (p_{ij}) with the conservative Q-matrix (q_{ij}), and $(r_{ij}^{(n)})=(r_{ij})^n$ as usual. Any d.p.M.C. with (r_{ij}) as transition matrix will be called a *jump chain* associated with the c.p. M.C. $\{x_t,\ t\in\boldsymbol{T}\}$.

The conditional distribution of ϱ_n relative to $\varrho_0,\ldots,\varrho_{n-1}$ or equivalently relative to $\tau_1,\ldots,\tau_n$ is the same as that relative to χ_n, and equals $e_{q_{\chi_n}}$ by Theorem 15.2. In particular, we have

$$\boldsymbol{E}\{\varrho_n|\varrho_0,\ldots,\varrho_{n-1}\}=q_{\chi_n}^{-1}$$

with probability one. Let $\varrho_n'=\min(\varrho_n,1)$. The sequence of random variables $\{\sigma_n,\ n\geqq 0\}$ where

$$\sigma_n=\sum_{\nu=0}^{n}\{\varrho_\nu'-\boldsymbol{E}[\varrho|\varrho_0,\ldots,\varrho_{\nu-1}]\} \tag{4}$$

with the Borel fields $\mathscr{F}\{\varrho_\nu,\ 0\leqq\nu\leqq n\}$ is a martingale. The summands on the right side of (4) are uniformly bounded by one; hence we can apply a theorem from martingale theoty (Doob [4; p. 323]) which asserts that the two series

$$\sum_n \varrho_n' \quad \text{and} \quad \sum_n \boldsymbol{E}[\varrho_n'|\varrho_0,\ldots,\varrho_{n-1}] \tag{5}$$

converge or diverge together with probability one. Clearly the first series in (5) and that in (2) converge or diverge together. On the other hand, a simple computation gives

$$\boldsymbol{E}[\varrho_n' \mid \varrho_1, \ldots, \varrho_{n-1}] = \boldsymbol{E}[\varrho_n' \mid \chi_n] = q_{\chi_n}^{-1}[1 - e^{-q_{\chi_n}}].$$

Now it is easy to see that for any positive sequence $\{q_n\}$, the two series $\sum_n q_n^{-1}[1-e^{-q_n}]$ and $\sum_n q_n^{-1}$ converge or diverge together. Hence the second series in (5) converge or diverge together with the series $\sum_n q_{\chi_n}^{-1}$. We have therefore proved the following theorem.

Theorem 1. For a.a. ω, $\tau_\infty(\omega) = \infty$ if and only if

$$\sum_n q_{\chi_n}^{-1}(\omega) = \infty. \tag{6}$$

Corollary 1. A sufficient condition that $\boldsymbol{P}\{\tau_\infty(\omega) < \infty \mid x_0(\omega) = i\} = 1$ is

$$\sum_{n=0}^{\infty} \sum_j r_{ij}^{(n)} q_j^{-1} < \infty. \tag{7}$$

Proof. The left member is simply the mathematical expectation of the left member of (6) given that $x_0(\omega) = i$. ▯

Simplest examples of this kind are where for each n there is a unique j_n such that $r_{ij_n}^{(n)} = 1$ and $\sum_n q_{j_n}^{-1} < \infty$; see Example 20.1. Turning to the other extreme, we have the next corollary.

Corollary 2. A sufficient condition that $\boldsymbol{P}\{\tau_\infty(\omega) = \infty\} = 1$, namely that almost all sample functions be step functions, or again that the minimal solution $(\overline{p}_{ij})$ be a transition matrix, is that the set of q_j be bounded above.

Simplest examples of this kind are the *finite case* where $\boldsymbol{I}$ is a finite set; and the *Poisson process* in which $\boldsymbol{I}$ is the set of nonnegative integers and $q_i = q_{i,i+1} = c$ where $0 < c < \infty$ for every $i \in \boldsymbol{I}$. The constant c is called the *parameter* of the Poisson process.

Another form of Corollary 2 is contained in the following theorem.

Theorem 2. The set $\{q_i, i \in \boldsymbol{I}\}$ is bounded above if and only if

$$\lim_{t \to 0} p_{ii}(t) = 1 \qquad \text{uniformly in } i \in \boldsymbol{I}; \tag{8}$$

or equivalently (2.E) holds uniformly in $i, j \in \boldsymbol{I}$. In this case the matrix (q_{ij}) is conservative.

Remark. In the second form the theorem remains true if the condition (1.B) is dropped.

Proof. If the q_i's are bounded above, then (8) holds by (3.2) with $i = j$ and $t = 0$. The converse will be proved by using the proof of Theorem 2.3. If (8) holds, then given $\varepsilon > 0$, there exists $t_0 = t_0(\varepsilon)$ such that

$$\inf_{i\in \boldsymbol{I}} \lim_{0\leqq t\leqq t_0} p_{ii}(t) > 1-\varepsilon .$$

This t_0 has the properties required of it in the cited proof. In particular if $nh < t_0$, then we have for every i and j:

$$p_{ij}(nh) \leqq \frac{p_{ij}(t_0)}{p_{ii}(t_0-nh)} < \frac{p_{ij}(t_0)}{1-\varepsilon};$$

and (2.13) yields for every $h < t_0/2$ and $n = [t_0/h]$:

$$\frac{p_{ij}(h)}{h} < \frac{2p_{ij}(t_0)}{(1-3\varepsilon)(1-\varepsilon)t_0}. \tag{9}$$

Summing over $j \neq i$ and letting $h \downarrow 0$, we obtain

$$q_i = \lim_{h\downarrow 0} \frac{1-p_{ii}(h)}{h} \leqq \frac{2}{(1-3\varepsilon)(1-\varepsilon)t_0}.$$

Next, the series $\sum_j p_{ij}(h)h^{-1}$ converges uniformly in $h \in (0, t_0/2)$ by (9); hence we may let $h \downarrow 0$ term by term below to obtain

$$0 = \lim_{h\downarrow 0} \left\{ \frac{1-p_{ii}(h)}{h} - \sum_{j\neq i} \frac{p_{ij}(h)}{h} \right\} = q_i - \sum_{j\neq i} q_{ij}. \quad \square$$

Instead of using Theorem 1, we can prove that the condition (8) implies almost all sample functions are step functions by using the argument in the proof of Theorem 5.7, applied now to all stable intervals. The probability that the set $\{x(\nu n^{-1}A, \omega), 0 \leqq \nu \leqq n\}$ contains $N+1$ different values does not exceed $\sup_{0\leqq t\leqq An^{-1}} [1-\inf_i p_{ii}(t)]^N$, and so on.

We now proceed to derive a necessary and sufficient condition for $\boldsymbol{P}\{\tau_\infty(\omega) = \infty \mid x_0(\omega) = i\} = 1$ or rather its opposite $\boldsymbol{P}\{\tau_\infty(\omega) < \infty \mid x_0(\omega) = i\} > 0$. This condition is a generalization of (7). Recall the definition of a sojourn set A (§ I.17) for the jump matrix (r_{ij}) and let $\Lambda = \{\omega\colon \chi_n(\omega) \in A \text{ for all } n \geqq 1\}$. Thus by definition we have $\boldsymbol{P}\{\Lambda \mid \chi_0(\omega) = i\} > 0$ for some $i \in A$. In the notation of taboo probability, if we write $A' = \boldsymbol{I} - A$, this may be written as

$$\lim_{n\to\infty} \sum_{j\in A} {}_{A'}r_{ij}^{(n)} > 0 \tag{10}$$

where

$${}_{A'}r_{ij}^{(n)} = \boldsymbol{P}\{\chi_\nu(\omega) \in A, 1 \leqq \nu < n; \chi_n(\omega) = j \mid \chi_0(\omega) = i\}.$$

Theorem 3. If there exists a sojourn set A for the jump matrix and an i satisfying both (10) and

$$\sum_{n=0}^{\infty} \sum_{j\in A} {}_{A'}r_{ij}^{(n)} q_j^{-1} < \infty, \tag{11}$$

then

$$\boldsymbol{P}\{\tau_\infty(\omega)<\infty \mid x_0(\omega)=i\}>0. \tag{12}$$

Conversely if (12) is true then there exists a sojourn set A and an i satisfying (10) and

$$\sum_{n=0}^{\infty}\sum_{j\in A} r_{ij}^{(n)}\, q_j^{-1}<\infty. \tag{13}$$

Remark. Since (11) is implied by (13) we may use the latter condition in both parts of the theorem. Furthermore we may allow some $q_j=0$ if we set the corresponding $q_j^{-1}=\infty$.

Proof. Let $\Delta_i=\{\omega\colon \chi_0(\omega)=i\}=\{\omega\colon x_0(\omega)=i\}$. As already remarked, the condition (10) is equivalent to $\boldsymbol{P}(\Lambda\mid\Delta_i)>0$. We have

$$\int_\Lambda q_{\chi_n(\omega)}^{-1}\boldsymbol{P}(d\omega\mid\Delta_i)=\sum_{j\in A}\boldsymbol{P}\{\Lambda;\chi_n(\omega)=j\mid\Delta_i\}q_j^{-1}\leqq\sum_{j\in A}{}_{A'}r_{ij}^{(n)}\, q_j^{-1};$$

consequently by (11)

$$\int_\Lambda\sum_n q_{\chi_n(\omega)}^{-1}\,\boldsymbol{P}(d\omega)\mid\Delta_i)<\infty.$$

Hence for a.a. ω in $\Delta_i\Lambda$, $\sum_n q_{\chi_n(\omega)}^{-1}<\infty$. It follows from Theorem 1 that $\boldsymbol{P}\{\tau_\infty(\omega)<\infty\mid\Delta_i\}\geqq\boldsymbol{P}(\Lambda\mid\Delta_i)>0$.

To prove the converse suppose that (12) is true; then there exists a $t>0$ and a $c>0$ such that

$$L_i(t)=\boldsymbol{P}\{\tau_\infty(\omega)\leqq t\mid\Delta_i\}=c.$$

Let $A=\{j\colon L_j(t)\geqq c\}$; we shall show that A is a sojourn set. Let

$${}_{A'}f_{ij}^{*}=\sum_{n=1}^{\infty}\boldsymbol{P}\{\chi_\nu(\omega)\in A,\ 1\leqq\nu<n;\ \chi_n(\omega)=j\mid\Delta_i\}.$$

If A is not a sojourn set, then by definition $\sum_{j\in A'}{}_{A'}f_{ij}^{*}=1$ and we have by the method of first entrance (§ I.4)

$$L_i(t)\leqq\sum_{j\in A'}{}_{A'}f_{ij}^{*}L_j(t).$$

Since $L_j(t)<c$ for every $j\in A'$ this implies that $L_i(t)<c$ which is a contradiction. Hence A is a sojourn set. Now define $\varrho_n^A(\omega)=\varrho_n(\omega)$ if $\chi_n(\omega)\in A$ and $=0$ otherwise; $\tau_\infty^A(\omega)=\sum_{n=0}^{\infty}\varrho_n^A(\omega)$. Then we have

$$\boldsymbol{E}\{\tau_\infty^A\mid\Delta_i\}=\sum_{n=0}^{\infty}\boldsymbol{E}\{\varrho_n^A\mid\Delta_i\}=\sum_{n=0}^{\infty}\sum_{j\in A}r_{ij}^{(n)}\,q_j^{-1}$$

which is the left member of (13). Define $\{\tau'_n, n \geqq 0\}$ inductively as follows: $\tau'_0 \equiv 0$; on the set $\{\omega: \tau'_n(\omega) + t < \tau_\infty(\omega)\}$ let

$$\tau'_{n+1}(\omega) = \inf\{s: s > \tau'_n(\omega) + t;\ x(s, \omega) \in A\}$$

if the s-set between the braces is not empty, otherwise $\tau'_{n+1}(\omega) = \infty$. We have $\boldsymbol{P}\{\tau'_1(\omega) < \infty \mid \varDelta_i\} \leqq 1 - L_i(t) = 1 - c$. Furthermore we have for each $n \geqq 0$, by the strong Markov property applied to the optional random variable τ'_n,

$$\begin{aligned} \boldsymbol{P}\{\tau'_{n+1}(\omega) < \infty \mid \tau'_n(\omega) < \infty\} &\leqq \sup_{j \in A} \boldsymbol{P}\{\tau'_n(\omega) + t < \tau_\infty(\omega) \mid x(\tau'_n(\omega), \omega) = j\} \\ &\leqq \sup_{j \in A}\{1 - L_j(t)\} \leqq 1 - c. \end{aligned}$$

Hence $\boldsymbol{P}\{\tau'_n(\omega) < \infty \mid \varDelta_i\} \leqq (1-c)^n$ and it is easy to see from the definitions that

$$\boldsymbol{E}(\tau^A_\infty) \leqq \sum_{n=0}^{\infty} \boldsymbol{P}\{\tau'_n(\omega) < \infty\}\, n t \leqq \sum_{n=0}^{\infty} (1-c)^n n t < \infty.$$

This proves (13). ▯

From the point of view of § 18, Theorem 3 is a criterion that the minimal solution of the two systems of differential equations for a given conservative Q-matrix be stochastic.

What does the sample function $x(t, \omega)$ do as t approaches the first infinity $\tau_\infty(\omega)$ (to be denoted by $\tau(\omega)$ below)? It is said to be reaching the boundary of a certain compactification of the state space $\boldsymbol{I}$. The question is how to achieve such a compactification so as to analyse the behavior of the sample functions by means of it. We shall be concerned with a finite $\tau(\omega)$, thus from the temporal rather than spatial point of view the purpose is to analyse the discontinuities of the sample functions which are not jumps, such as the τ here.

We begin with the observation that for the minimal transition matrix $\bar{\Pi} = (\bar{p}_{ij})$ completed by ϑ, notions of recurrence as defined in § 10 apply as well but are distinct from those for the given transition matrix $\Pi = (p_{ij})$. In particular a state which is recurrent with respect to $\bar{\Pi}$ is recurrent with respect to Π, but the converse is false. Now if i is $\bar{\Pi}$-recurrent then by (10.6) applied to the minimal chain, under the hypothesis that $x(0, \omega) = i$ the set $S_i(\omega) \cap (0, \tau(\omega))$ is unbounded for a.e. ω; hence $\tau(\omega) = \infty$ and $L_i(t) \equiv 0$. This means, starting from i the probability is zero that the first infinity be finite, or that the boundary be reached in finite time. Such a state would play only a nuisance role in the discussion below and will be excluded for the sake of simplicity. Thus we shall assume that every i in $\boldsymbol{I}$ is $\bar{\Pi}$-nonrecurrent. It follows then from

Theorem 10.4 and its Corollary 2 that if we put for every i and j in $\boldsymbol{I}$:

$$\zeta_{ij} = \int_0^\infty \bar{p}_{ij}(t)\,dt, \qquad \zeta_j = \sum_i p_i \zeta_{ij} = \int_0^\infty \bar{p}_j(t)\,dt,$$

where $\{p_i\}$ is the initial distribution of the given M.C., and $\{\bar{p}_i(t)\}$ its absolute distribution at time t, then we have $0 \leqq \zeta_{ij} < \infty$ and $0 < \zeta_j < \infty$ ($0 < \zeta_j$ since $\boldsymbol{I}$ is minimial).

We shall now actually turn the M.C., or equivalently the corresponding minimal chain, from time τ backwards. That is, we consider the process $x(\tau(\omega) - t, \omega)$ for $0 < t \leqq \tau(\omega) < \infty$. The basic result is as follows.

Theorem 4. For $0 < t_1 < \cdots < t_m < \infty$ and arbitrary $j_1, \ldots, j_m$ in $\boldsymbol{I}$, we have

$$(14) \qquad \begin{cases} \boldsymbol{P}\{t_m \leqq \tau(\omega) < \infty;\ x(\tau(\omega) - t_n, \omega) = j_n, \quad 1 \leqq n \leqq m\} \\ = \zeta_{j_m} \prod\limits_{n=m}^{2} \bar{p}_{j_n j_{n-1}}(t_n - t_{n-1})\, l_{j_1}(t_1) \end{cases}$$

where $l_j(\cdot) = L_j'(\cdot)$ is given in Theorem 18.6.

Proof. Let $t_m = n_0 \delta$ where n_0 is a positive integer and $\delta > 0$; and put

$$a_n = \boldsymbol{P}\{n\delta \leqq \tau(\omega) < (n+1)\delta;\ x(\tau(\omega) - t_n, \omega) = j_n, 1 \leqq n \leqq m\}.$$

Since every state is stable, we have byTheorem 15.2 and (17.16)

$$(15) \qquad \begin{cases} a_n \geqq \boldsymbol{P}\{n\delta \leqq \tau(\omega) < (n+1)\delta;\ x(u - t_n, \omega) = j_n \\ \qquad\qquad \text{for} \quad n\delta \leqq u \leqq (n+1)\delta;\ 1 \leqq n \leqq m\} \\ = \bar{p}_{j_m}(n\delta - t_m)\, e^{-q_{j_m}\delta} \prod\limits_{n=m}^{2} \bar{p}_{j_n j_{n-1}}(t_n - t_{n-1} - \delta)\, e^{-q_{j_{n-1}}\delta} \times \\ \qquad \times \boldsymbol{P}\{n\delta \leqq \tau(\omega) < (n+1)\delta \mid x((n+1)\delta - t_1, \omega) = j_1\}. \end{cases}$$

The last-written factor above is equal to

$$L_{j_1}(t_1) - L_{j_1}(t_1 - \delta) = l_{j_1}(t_1 - \theta\delta)\,\delta$$

where $0 \leqq \theta \leqq 1$. Since $\bar{p}_{j_m}(\cdot)$ is continuous and integrable over $(0, \infty)$, we have certainly

$$\lim_{\delta \downarrow 0} \sum_{n=n_0}^{\infty} \bar{p}_{j_m}(n\delta - t_m)\,\delta \geqq \int_0^\infty \bar{p}_{j_m}(u)\,du = \zeta_{j_m};$$

in fact equality holds above. Thus if we sum (15) over $n \geqq n_0$, then let $\delta \downarrow 0$, and recall that $l_{j_1}(\cdot)$ is continuous we obtain (14) with "$\geqq$" replacing the "$=$" there. Now if we sum over all $j_1, \ldots, j_m$ and use the crucial equation (18.11), we have

$$\begin{aligned}\sum_{j_1}\cdots\sum_{j_m}\boldsymbol{P}\{t_m\leqq\tau(\omega)<\infty;\ x(\tau(\omega)-t_n,\omega)=j_n, 1\leqq n\leqq m\}&\geqq\sum_{j_m}\zeta_{j_m}\,l_{j_m}(t_m)\\ =\sum_j\int_0^\infty\sum_i p_i\,\overline{p}_{ij}(s)\,l_j(t_m)\,ds&\\ =\sum_i p_i\int_0^\infty l_i(s+t_m)\,ds=\sum_i p_i[L_i(\infty)-L_i(t_m)]&.\end{aligned}$$

The first member above obviously does not exceed $\boldsymbol{P}\{t_m\leqq\tau(\omega)<\infty\}$ which is equal to the last member by (3). Hence there is equality above which implies the same in (14). □

Corollary 1. For $t>0$ and $j\in\boldsymbol{I}$, we have

$$\boldsymbol{P}\{t\leqq\tau(\omega)<\infty;\ x(\tau(\omega)-t,\omega)=j\}=\zeta_j\,l_j(t).$$

Corollary 2. For $t>0$ and i, j in $\boldsymbol{I}$, we have

$$\boldsymbol{P}\{x(0,\omega)=i\mid t\leqq\tau(\omega)<\infty;\ x(\tau(\omega)-t,\omega)=j\}=\frac{p_i\,\zeta_{ij}}{\zeta_j}.$$

We can now give a formal definition as follows. *The reverse chain in* $[0,\tau(\omega))$ is the M.C. $\{y_t, t\in\boldsymbol{T}^0\}$ where

$$y_t(\omega)=y(t,\omega)=\begin{cases}x(\tau(\omega)-t,\omega) & \text{if } t\leqq\tau(\omega);\\ \vartheta & \text{if } t>\tau(\omega).\end{cases}$$

If we put for every i and j in $\boldsymbol{I}$:

$$(16)\qquad\begin{cases}\psi_{ij}(t)=\dfrac{\zeta_j\,\overline{p}_{ji}(t)}{\zeta_i}, & \psi_{i\vartheta}(t)=1-\sum\limits_j\psi_{ij}(t);\\ \psi_j(t)=\zeta_j l_j(t), & \psi_\vartheta(t)=1-\sum\limits_j\psi_j(t);\end{cases}$$

then Theorem 4 may be restated as follows.

Theorem 4′. The reverse chain $\{y_t, t\in\boldsymbol{T}^0\}$ in $(0,\tau(\omega))$ is an open M.C., with the transition matrix (ψ_{ij}), $i, j\in\boldsymbol{I}\cup\{\vartheta\}$; and the absolute distribution $\{\psi_j(t)\}$, $j\in\boldsymbol{I}\cup\{\vartheta\}$, at time t.

The main limit theorem for the approach to the first infinity is obtained by applying the martingale convergence theorem to the reverse chain. Observe that $\psi_{y(s,\omega),i}(t-s)$ is a version of the conditional probability

$$\boldsymbol{P}\{y(t,\omega)=i\mid y(r,\omega), 0<r\leqq s\}=\boldsymbol{P}\{y(t,\omega)=i\mid y(s,\omega)\};$$

cf. the proof of Theorem 6.1. The left member above forms a martingale for decreasing s, hence a separable version $\{z_s, 0<s\leqq t\}$ of it has the

property that $\lim_{s \downarrow 0} z(s, \omega)$ exists for a.e. ω (Doob [4, Theorem VII.11.2]). If R is a separability set, then for each fixed t there exists a set Ω_t with $\boldsymbol{P}(\Omega_t) = 1$ such that for every $\omega \in \Omega_t$ we have

$$(17) \qquad \lim_{r \in R,\, r \downarrow 0} \psi_{y(r,\omega),i}(t-r) = \lim_{r \in R,\, r \downarrow 0} z(r, \omega) = z(0+, \omega).$$

Theorem 5. For each t, the following limit exists for every i in $\boldsymbol{I}$ and a.e. ω:

$$(18) \qquad \lim_{s \downarrow 0} \frac{p_{i,\, x(\tau(\omega)-s,\, \omega)}(t-s)}{\zeta_{x(\tau(\omega)-s,\, \omega)}} = z(0+, \omega).$$

Proof. Observing that $\tau(\omega) > 0$ a.e., this is just (17) if s is restricted to R. Since all states are stable, it follows from Theorem 5.7 that for a.e. ω, and every s, $y(s, \omega) = x(\tau(\omega) - s, \omega)$ is a constant, say k, in an interval containing s (possibly as an endpoint). This fact and the continuity of $\bar{p}_{ik}(\cdot)$ imply that $\bar{p}_{i,\, y(s,\omega)}(t-s)/\zeta_{y(s,\omega)}$ is equal to the limit of $\bar{p}_{i,\, y(r,\omega)}(t-r)/\zeta_{y(r,\omega)}$ as $r \to s$ on R from at least one side of s. Hence the limit in (18) exists and is the same as in (17). ▯

A similar argument beginning with Corollary 2 to Theorem 4 yields the existence of the following limit for every i in $\boldsymbol{I}$ and a.e. ω:

$$\lim_{s \downarrow 0} \frac{\zeta_{i,x(\tau(\omega)-s,\, \omega)}}{\zeta_{x(\tau(\omega)-s,\, \omega)}}.$$

Adopting a notion due to R. S. Martin, a sequence of states $\{j_\nu\}$ tending to ∞ as integers is said to *converge to a boundary point* iff for each i, the numeral sequence $\{\zeta_{ij_\nu}/\zeta_{j_\nu}\}$ converges (finitely). Two such sequences are said to converge to the same boundary point iff the two corresponding numerical sequences have the same limit. With this definition it follows from the above that as $s \downarrow 0$, $x(\tau(\omega) - s, \omega)$ converges to a boundary point $x(\tau(\omega) - 0, \omega)$ which is a random variable taking values in a boundary set $\boldsymbol{B}$. The set $\boldsymbol{I} \cup \boldsymbol{B}$ may be made into a compact metric space in which $\boldsymbol{I}$ is dense. Thus the result is a certain compactification of the minimal state space of an arbitrary minimal chain. The Π-recurrent states which have been excluded in the above discussion may be added to the compactified space in a trivial way.

What lies beyond the first infinity? A satisfactory answer must make use of some such compactification as alluded to above, and should yield an analysis of the process in terms of the "boundary" set up for the purpose. This has been done quite recently in the case of a finite number of essentially different ways of "going to infinity" (or "reaching the boundary"), the so-called "exits", but we cannot enter into it here. In what follows we shall merely illustrate the simplest possibilities of "coming back from infinity", the so-called "entrances".

Let us assume that $x_+(\tau_\infty(\omega),\omega)\in\boldsymbol{I}$ with probability one. The strong Markov property then holds and we may pass on to the post-α process. The latter has a first infinity which is the second infinity of the original process, and so on. To study the variety of possibilities we may pose the following problem. Given the substochastic transition matrix $(\bar{p}_{ij})$ which is the minimal solution to the two systems (2) and (3) of § 18 associated with the Q-matrix (q_{ij}), to construct a M.C. $\{x_t, t\in\boldsymbol{T}\}$ for which (17.16) is satisfied. This problem has a trivial solution if we are allowed to enlarge $\boldsymbol{I}$ (cf. Example 20.1); indeed the simplest solution is then a minimal chain defined after Theorem 18.3. But a less trivial solution will now be given.

Let

$$L=\sum_{i\in\boldsymbol{I}}p_iL_i,\qquad L^{0*}=\epsilon,\qquad L^{(n+1)*}=L^{n*}*L,$$

$$K_i=\sum_{n=0}^{\infty}(L_i*L^{n*});$$

so that L is the distribution of τ_∞; for the interpretation of K_i compare (16.5). We may now state the following construction which gives a special solution to the problem posed.

Theorem 6. Let $\{x_t^{(n)}, t\in\boldsymbol{T}\}$, $n\geqq 1$, be a sequence of mutually independent minimal chains having the same initial distribution and corresponding to a given conservative Q-matrix with index set $\boldsymbol{I}$. Let $\tau^{(n)}$ be the first infinity of $\{x_t^{(n)}, t\in\boldsymbol{T}\}$. Define $\tau^{(0)}=0$, $\sigma^{(n)}=\sum_{\nu=0}^{n}\tau^{(\nu)}$; and

$$x(t,\omega)=x^{(n)}\big(t-\sigma^{(n-1)}(\omega),\omega\big)\quad\text{if}\quad\sigma^{(n-1)}(\omega)\leqq t<\sigma^{(n)}(\omega).$$

Then $\{x_t, t\in\boldsymbol{T}\}$ is a M.C. with the minimal state space $\boldsymbol{I}$ (no adjoined state is necessary), the initial distribution $\{p_i, i\in\boldsymbol{I}\}$ and the transition matrix (p_{ij}) where

$$p_{ij}(t)=\bar{p}_{ij}(t)+\int_0^t\bar{p}_j(t-s)\,dK_i(s). \tag{19}$$

Proof. If $\tau^{(1)}=\infty$ with probability one then $(\bar{p}_{ij})$ is a transition matrix, $p_{ij}=\bar{p}_{ij}$ and $x(t,\omega)\equiv x^{(1)}(t,\omega)$ so that there is nothing to prove. We may therefore suppose that each $\tau^{(n)}$ is finite with positive probability. Since $\lim_{n\to\infty}\sigma^{(n)}=\infty$ with probability one the process $\{x_t, t\in\boldsymbol{T}\}$ is well-defined.

Let $0=t_0<t_1<\cdots<t_{l+1}$; $j_1,\ldots,j_{l+1}\in\boldsymbol{I}$; we write for the sake of brevity

$$\Lambda_{kl}=\bigcap_{\nu=k}^{l}\{\omega: x(t_\nu,\omega)=j_\nu\},\qquad\Lambda_0=\Omega,\qquad\Lambda_l=\Lambda_{1l};$$

$$\boldsymbol{P}_\Lambda(\cdot)=\boldsymbol{P}(\cdot\,|\,\Lambda),\qquad\boldsymbol{E}_\Lambda(\cdot)=\boldsymbol{E}(\cdot\,|\,\Lambda).$$

Lemma. We have for each $l\geqq 0$ and $m\geqq 0$,

$$\left\{\begin{aligned}&\boldsymbol{P}\{\Lambda_{l+1};\,\sigma^{(m)}(\omega)\leqq t_l<t_{l+1}<\sigma^{(m+1)}(\omega)\}\\&\qquad=\boldsymbol{P}\{\Lambda_l;\,\sigma^{(m)}(\omega)\leqq t_l<t_{l+1}<\sigma^{(m+1)}(\omega)\}\,\bar{p}_{j_l j_{l+1}}(t_{l+1}-t_l)\end{aligned}\right.\tag{20}$$

and if $t_l<u$,

$$\left\{\begin{aligned}&\boldsymbol{P}\{\Lambda_l;\,\sigma^{(m)}(\omega)\leqq t_l<\sigma^{(m+1)}(\omega);\,\sigma^{(m+1)}(\omega)\leqq u\}\\&\qquad=\boldsymbol{P}\{\Lambda_l;\,\sigma^{(m)}(\omega)\leqq t_l<\sigma^{(m+1)}(\omega)\}\,L_{j_l}(u-t_l)\,.\end{aligned}\right.\tag{21}$$

Proof of the lemma. If we set $\Lambda_{k-1}^{(m)}=\Lambda_{k-1}\cap\{\omega: t_{k-1}<\sigma^{(m)}(\omega)\}$, the left member of (15) is equal to

$$\begin{aligned}&\sum_{k=1}^{l}\boldsymbol{P}\{\Lambda_{k-1};\,t_{k-1}<\sigma^{(m)}(\omega)\leqq t_k<t_{l+1}<\sigma^{(m+1)}(\omega);\,\Lambda_{k,l+1}\}\\&\quad=\sum_{k=1}^{l}\boldsymbol{P}\{\Lambda_{k-1}^{(m)}\}\boldsymbol{E}_{\Lambda_{k-1}^{(m)}}[\boldsymbol{P}_{\Lambda_{k-1}^{(m)}}\{\Lambda_{k,l+1};\,\sigma^{(m)}(\omega)\leqq t_k<t_{l+1}<\sigma^{(m+1)}(\omega)\,|\,\sigma^{(m)}\}]\\&\quad=\sum_{k=1}^{l}\boldsymbol{P}\{\Lambda_{k-1}^{(m)}\}\boldsymbol{E}_{\Lambda_{k-1}^{(m)}}[\boldsymbol{P}_{\Lambda_{k-1}^{(m)}}\{x^{(m+1)}(t_\nu-\sigma^{(m)}(\omega))=j_\nu,\ k\leqq\nu\leqq l+1;\\&\qquad\qquad\qquad\sigma^{(m)}(\omega)\leqq t_k<t_{l+1}<\sigma^{(m+1)}(\omega)\,|\,\sigma^{(m)}\}]\,.\end{aligned}$$

Since $\{x_t^{(m+1)},\,t\in\boldsymbol{T}\}$ is independent of $\Lambda_{k-1}^{(m)}$, we can replace the $\boldsymbol{P}_{\Lambda_{k-1}^{(m)}}$ above by $\boldsymbol{P}$; we have then, using the Markov property of $\{x_t^{(m+1)}\}$,

$$\begin{aligned}&\sum_{k=0}^{l}\boldsymbol{P}\{\Lambda_{k-1}^{(m)}\}\boldsymbol{E}_{\Lambda_{k-1}^{(m)}}[\boldsymbol{P}\{x^{(m+1)}(t_\nu-\sigma^{(m)}(\omega))=j_\nu,\ k\leqq\nu\leqq l;\\&\qquad\qquad\sigma^{(m)}(\omega)\leqq t_k<t_{l+1}<\sigma^{(m+1)}(\omega)\,|\,\sigma^{(m)}\}\,\bar{p}_{j_l j_{l+1}}(t_{l+1}-t_l)]\,.\end{aligned}$$

Retracing our steps, this reduces to the right member of (20). The proof of (21) is similar. ▯

Proof of Theorem 6. We have

$$\begin{aligned}\boldsymbol{P}\{\Lambda_{l+1}\}=&\sum_{1\leqq m\leqq n\leqq<\infty}\boldsymbol{P}\{\Lambda_l;\,\sigma^{(m-1)}(\omega)\leqq t_l<\sigma^{(m)}(\omega);\,\sigma^{(n-1)}(\omega)\leqq t_{l+1}\\&<\sigma^{(n)}(\omega);\,x(t_{l+1},\omega)=j_{l+1}\}=\sum_{1\leqq m=n<\infty}+\sum_{1\leqq m<n<\infty}\,.\end{aligned}$$

By the Lemma, the first sum is equal to

$$\begin{aligned}&\sum_{m=1}^{\infty}\boldsymbol{P}\{\Lambda_l;\,\sigma^{(m-1)}(\omega)\leqq t_l<\sigma^{(m)}(\omega)\}\,\bar{p}_{j_l j_{l+1}}(t_{l+1}-t_l)\\&\qquad\qquad=\boldsymbol{P}(\Lambda_l)\,\bar{p}_{j_l j_{l+1}}(t_{l+1}-t_l)\,.\end{aligned}$$

To evaluate the second sum we note that if $m<n$ and $u<v$,

$$\begin{aligned}&\boldsymbol{P}\{\sigma^{(n-1)}(\omega)\leqq v\,|\,\sigma^{(m)}=u\}=L^{(n-1-m)*}(v-u)\,;\\&\boldsymbol{P}\{x(t_{l+1},\omega)=j_{l+1},\,\sigma^{(n-1)}(\omega)\leqq t_{l+1}<\sigma^{(n)}(\omega)\,|\,\sigma^{(n-1)}=v\}\\&\qquad=\boldsymbol{P}\{x^{(n)}(t_{l+1}-\sigma^{(n-1)}(\omega))=j_{l+1}\,|\,\sigma^{(n-1)}=v\}\\&\qquad=\bar{p}_{j_{l+1}}(t_{l+1}-v)\,,\end{aligned}$$

for certain versions of the conditional probabilities. Therefore if $m<n$, we have by the Lemma,

$$\begin{aligned} P_{mn} &= \boldsymbol{P}\{\Lambda_l;\, \sigma^{(m-1)}(\omega) \leqq t_l < \sigma^{(m)}(\omega); \\ &\qquad \sigma^{(n-1)}(\omega) \leqq t_{l+1} < \sigma^{(n)}(\omega);\, x(t_{l+1}, \omega) = j_{l+1}\} \\ &= \int_{t_l}^{t_{l+1}} \boldsymbol{P}\{\Lambda_l;\, \sigma^{(m-1)}(\omega) \leqq t_l < \sigma^{(m)}(\omega)\} d_u L_{j_l}(u-t_l) \times \\ &\quad \times \int_{u}^{t_{l+1}} \overline{p}_{j_{l+1}}(t_{l+1}-v)\, d_v L^{(n-1-m)*}(v); \end{aligned}$$

$$\begin{aligned} \sum_{m=1}^{\infty} \sum_{n=m+1}^{\infty} P_{mn} &= \sum_{m=1}^{\infty} \boldsymbol{P}\{\Lambda_l;\, \sigma^{(m-1)}(\omega) \leqq t < \sigma^{(m)}(\omega)\} \times \\ &\quad \times \int_{u=t_l}^{t_{l+1}} \int_{v=u}^{t_{l+1}} \overline{p}_{j_{l+1}}(t_{l+1}-v)\, d_u L_{j_l}(u-t_l)\, d_v \left\{ \sum_{n=0}^{\infty} L^{n*}(v-u) \right\} \\ &= \boldsymbol{P}(\Lambda_l) \int_{t_l}^{t_{l+1}} \overline{p}_{j_{l+1}}(t_{l+1}-v)\, d_v \left\{ \sum_{n=0}^{\infty} [L_{j_l} * L^{n*}](v-t_l) \right\} \\ &= \boldsymbol{P}(\Lambda_l) \int_{0}^{t_{l+1}-t_l} \overline{p}_{j_{l+1}}(t_{l+1}-t_l-v)\, dK_{j_l}(v). \end{aligned}$$

Combining the two sums, we obtain

$$\begin{aligned} \boldsymbol{P}(\Lambda_{l+1}) &= \boldsymbol{P}(\Lambda_l) \left\{ \overline{p}_{j_l j_{l+1}}(t_{l+1}-t_l) + \int_{0}^{t_{l+1}-t_l} \overline{p}_{j_{l+1}}(t_{l+1}-t_l-v)\, dK_{j_l}(v) \right\} \\ &= \boldsymbol{P}(\Lambda_l)\, p_{j_l j_{l+1}}(t_{l+1}-t_l) \end{aligned}$$

in view of (14). This proves the theorem by induction on l. ▯

It can be directly verified that (p_{ij}) as given in (19) is a transition matrix by using Laplace transforms and the so-called *resolvent equation* (see FELLER [5]). A slight generalization of the construction may be obtained by allowing the initial distribution of $\{x_t^{(1)}\}$ to be different from the common one of $\{x_t^{(n)}\}$, $n \geqq 2$. In any case the resulting M.C. has sample functions whose sets of discontinuities (cf. Corollary to Theorem 7.5) are well-ordered with probability one. Clearly $p'_{ij}(0) = q_{ij}$ and it follows from Theorem 17.4 that the matrix (p_{ij}) is a solution to the first but not the second system of differential equations, provided that $(\overline{p}_{ij})$ itself is not a transition matrix. Since the choice of $\{p_i, i \in \boldsymbol{I}\}$ is arbitrary we have the following corollary which furnishes a converse to Theorem 18.3.

Corollary. For a given conservative Q-matrix, if the minimal solution $(\overline{p}_{ij})$ is not a transition matrix, then there are infinitely many transition matrices which are solutions to the first but not the second system of differential equations. In particular, if the minimal solution is not the unique solution to the first system then there are infinitely many solutions.

A new type of construction beyond the first infinity is possible if we assume the existence of a simple kind of "entrance law" as follows. Let $\{\eta_i, i \in \boldsymbol{I}\}$ be functions on $\boldsymbol{T}^0$ satisfying the conditions, for $s>0$ and $t \geqq 0$:

$$(22)\qquad \begin{cases} \eta_j(t) \geqq 0; \\ \eta_j(s+t) = \sum_i \eta_i(s) \bar{p}_{ij}(t); \\ \lim_{t \downarrow 0} \sum_j \eta_j(t) = 1. \end{cases}$$

Observe that the second condition in (22) implies that

$$\sum_j \eta_j(s+t) = \sum_i \eta_i(s)\left[1 - L_i(t)\right]$$

so that $\sum_j \eta_j(t)$ is nonincreasing in t and the limit in the third condition always exists but may be infinite. The latter case yields a different kind of entrance law not envisaged here. Note also that the $\{\bar{p}_i(\cdot)\}$ in Theorem 6 satisfies these conditions, indeed even for $s=0$ in the second condition and without the limit in the third. In general the existence of $\eta_j(0) = \lim_{t \downarrow 0} \eta_j(t)$ is a consequence of Theorem 2.3, but $\sum_j \eta_j(0)$ may be less than $\lim_{t \downarrow 0} \sum_j \eta_j(t)$. This is the essential difference between Theorem 6 and the present more general case. Now instead of a closed minimal chain we can construct an open one $\{u_t, t \in \boldsymbol{T}^0\}$ as discussed in § 4. This has minimal state space $\boldsymbol{I} \cup \{\vartheta\}$ and the transition matrix $(\bar{p}_{ij})$ completed by ϑ as in Theorem 3.3. Its absolute distribution at $t>0$ is given by $\eta_j(t)$, $j \in \boldsymbol{I} \cup \{\vartheta\}$, where $\eta_\vartheta(t) = 1 - \sum_{i \in \boldsymbol{I}} \eta_i(t)$. Specifically, for $0 < t_0 < t_1 < \cdots < t_l < t_{l+1}$, and arbitrary $i_1, \ldots, i_l$ in $\boldsymbol{I}$, we have

$$\boldsymbol{P}\{u(t_\nu, \omega) = i_\nu, 1 \leqq \nu \leqq l; u(t_{l+1}, \omega) = \delta\}$$
$$= \sum_{i_0 \in \boldsymbol{I}} \eta_{i_0}(t_0) \prod_{\nu=1}^{l} \bar{p}_{i_{\nu-1}, i_\nu}(t_\nu - t_{\nu-1}) L_{i_l}(t_{l+1} - t_l).$$

For each $t_0 > 0$, the process $\{u_t, t_0 \leqq t < \infty\}$ is a closed minimal chain, hence in each such interval $[t_0, \infty)$, a.e. sample function $u(\cdot, \omega)$ is a step function with an infinite number of jumps ending in ϑ, absorbing states being excluded. But the points of jump in $(0, t_0)$ is either a finite set or an infinite countable set of the ordinal type of the negative integers in their algebraic order. Finally let $\{u_t^{(n)}\}$, $n \geqq 1$, be a sequence of mutually independent open M.C.'s just described with the same given entrance law $\{\eta_j(\cdot)\}$. Each of them has a "first infinity" $\tau^{(n)}$ which is just the entrance time into ϑ. Now the construction given in Theorem 6 may be carried out with the $x_t^{(n)}$'s replaced by the $u_t^{(n)}$'s. Equation (19) becomes

$$p_{ij}(t)=\bar{p}_{ij}(t)+\int_0^t \eta_j(t-s)\,dK_i(s)\,. \tag{23}$$

[Analytic verification that this (p_{ij}) is a stochastic transition matrix is not difficult.] As $t\downarrow\sigma^{(n)}(\omega)$ defined as in Theorem 6, $x(t,\omega)$ tends to each j with probability $\eta_j(0)$ and to ∞ with probability $1-\sum_j \eta_j(0)$. In analogy with Theorem 15.6, this corresponds to the phenomenon of exit from the boundary and re-entry into $\boldsymbol{I}$. The second system of equations of § 17 will be satisfied if and only if $\sum_j \eta_j(0)=0$, by Theorem 17.5.

The construction described above is only a very special kind of (p_{ij}) that can in general be obtained from a given $(\bar{p}_{ij})$ or (q_{ij}). The rest is part of the boundary theory, an introduction to which is given in CHUNG [12] and [14], where further references to the literature on the subject may be found.

Notes. The proof of Theorem 1 seems new; that of Theorem 2 is due to KENDALL (see LOÈVE [1; p. 590]); for the method sketched after Theorem 2 applied in a more general case see DOEBLIN [3]. Theorem 3 is due to FELLER [1] whose proof is couched in analytical terms; the present proof, attributed to DYNKIN, follows DOBRUŠIN [2] who deduced Theorem 1 from it.

Theorems 4 and 5 are from CHUNG [11]; they are inspired by HUNT'S [1] handling of a result of DOOB [5]. The latter corresponds to the discrete parameter version of Thoerem 5, and may be deduced from it; see the last paragraph of CHUNG [11], where however for "$\sum_i \psi_i^{-1}<\infty$" read "$\sum_i \gamma_i \sum_n \sum_j \psi_{ij}^{(n)}\, \psi_j^{-1}<\infty$". DOOB and HUNT are interested in a potential theory rather than the analysis of a M.C. in continuous time. From the point of view of this monograph, the test of a boundary lies in its *crossing*. It turns out that if the Martin boundary is countable so that the distribution of $x(\tau(\omega)-0,\omega)$ mentioned after Theorem 5 is atomic, then each atom corresponds to an atomic almost closed set as defined in § I.17. In such a case the crucial crossing may be done by direct methods (see § 4 of CHUNG [12]) and whatever boundary is set up serves merely as a prop. For literature on the boundary theory for M.C.'s see FELLER [4], [5]; REUTER [3], [4]; NEVEU [2], [3], DAVID WILLIAMS [2], [3], and CHUNG [12], [14]. The last two papers contain an analysis of the structure of all M.C.'s with stable states and a conservative Q-matrix, whose "passable boundary" consists of a finite number of "distinguishable" atoms. This includes a solution of the so-called "complete construction problem", initiated by FELLER [5]; namely to find all transition matrices of those M.C.'s just specified. Inklings of such constructions are given in Theorem 6 and the subsequent discussion, but the general case is rather more complicated. Above all, the study of the behavior of the process in relation to its boundary is carried out in these two papers.

Theorem 6 is a reformulation of a more direct construction given by DOOB [2] without recourse to the properties of $(\bar{p}_{ij})$ and requiring transfinite induction which proves to be unnecessary. The proof given here seems new and shows that actually only a sequence of $\sigma^{(n)}$, the so-called "blow-ups", is needed. DOOB's construction can also be verified by semigroup theory; see REUTER [2]. For further results on the construction problem see REUTER [2], [3], FELLER [5] and DORBUŠIN [4].

§ 20. Examples

In this section we give a number of examples to illustrate the various possibilities of sample function behavior. They imply that certain general propositions in the preceding sections are not vacuous; for instance the existence of an instantaneous state, the indispensability of the adjoined state, etc. Such examples have been erstwhile referred to as "pathological"; in truth they represent the "normal situation" in the general context of this book.

It is well known that we can construct a probability triple on which exist denumerably many independent random variables with given distributions, or M.C.'s with given transition matrices. We need not be specific about the triple in what follows. The letters i, j, k denote non-negative integers which serve as states as well as indices; we shall also allow states to be denoted by certain symbols which are not specified to be real numbers.

We begin with some general considerations. Let τ be a random variable on Ω which is positive but may be infinite. For each $t \in \boldsymbol{T}$ (or $\boldsymbol{T}^0$) let y_t be a random variable defined on the set $\{\omega: t < \tau(\omega)\}$. The collection $\{y(t, \omega), t < \tau(\omega)\}$ or more simply $\{y_t, t < \tau\}$ will be called a *stochastic process stopped at* τ, and τ the *stopping random variable*. If $\tau = \infty$ a.e. this becomes an ordinary process. Suppose now that the stopped process is denumerably-valued with $\boldsymbol{I}$ as its minimal state space. Then the defining equation (4.1) of the Markov property is still meaningful for the stopped process. We shall call $\{y_t, t < \tau\}$ a *M.C. stopped at* τ iff (4.1) and (4.2) hold when the x's there are replaced by the y's; and we shall call the substochastic transition matrix (p_{ij}) its matrix. Let

$\vartheta \notin \bar{\boldsymbol{I}}$, $\hat{\boldsymbol{I}} = \boldsymbol{I} \cup \{\vartheta\}$, and define $\{\hat{y}_t, t \in \boldsymbol{T}$ (or $\boldsymbol{T}^0$)$\}$ as follows:

$$(1) \qquad \begin{cases} \hat{y}(t, \omega) = y(t, \omega) & \text{if } t < \tau(\omega); \\ \hat{y}(t, \omega) = \vartheta & \text{if } \tau(\omega) \leqq t < \infty. \end{cases}$$

It is obvious that $\{y_t, t < \tau\}$ is a stopped M.C. if and only if $\{\hat{y}_t, t \in \boldsymbol{T}$ (or $\boldsymbol{T}^0$)$\}$ is a M.C. The latter will be called the *minimal completion* of the former. The transition matrix of the minimal completion is the $(\hat{p}_{ij})$ given by Theorem 3.3, unless $\tau = \infty$ a.e. in which case it is (p_{ij}). Furthermore, given any substochastic transition matrix (p_{ij}) there exists a stopped M.C. $\{y_t, t < \tau\}$ whose matrix coincides with it. The conditional distribution of τ is given as follows:

$$L_i(t) = \boldsymbol{P}\{\tau(\omega) \leqq t \mid y(0, \omega) = i\} = 1 - p_{i\boldsymbol{I}}(t).$$

For any $s \geqq 0$, $t > 0$ and $\Lambda \in \mathscr{F}\{y_u, u \leqq s\}$ we have

$$(2) \qquad L_i(t) = \boldsymbol{P}\{\tau(\omega) \leqq s + t \mid \Lambda; y(s, \omega) = i\}.$$

This follows from the Markov property of the stopped process since the right member above is the conditional probability of $y(s+t, \omega)$ being undefined or in other words $\notin \boldsymbol{I}$. If (p_{ij}) is substochastic then $L_i(t) = \hat{p}_{i\delta}(t)$ and is continuous by Theorems 1.1 and 3.3.

Examples of stopped M.C. are given below.

(a) The minimal chain discussed in §§ 18—19, where the stopping random variable τ is the first infinity τ_∞ and the matrix is the minimal solution $(\bar{p}_{ij})$.

(b) The M.C. stopped at the first exit from a nonabsorbing stable state k, discussed in § 15, where τ is $\tau_1'(k)$ and the matrix elements are given in the notation of § 11 as follows:

$$p_{ij}(t) = {}_kp_{ij}(t) \quad \text{if } i \neq k, j \neq k;$$

$$p_{ik}(t) = \int_0^t e^{-q_k(t-s)}\, dF_{ik}(s), \qquad p_{ki}(t) \equiv 0, \qquad \text{if } i \neq k; \qquad p_{kk}(t) \equiv e^{-q_k t}.$$

(c) The M.C. stopped at the first entrance into the taboo set H, discussed in § 11, with $\tau = \inf\limits_{k \in H} \alpha_k$ where α_k is the first entrance time into the state k, and with the matrix $({}_Hp_{ij})$, $i, j \in \boldsymbol{I} - H$.

Two stopped processes $\{y_t, t < \tau\}$ and $\{z_t, t < \tau'\}$ are said to be mutually independent if the two collections $\{\tau, y_t\}$ and $\{\tau', z_t\}$ are mutually independent.

Suppose furthermore that the two index sets $\boldsymbol{I}$ and $\boldsymbol{I}'$ are disjoint and the two matrices are standard. We then define a process $\{x_t, t \in \boldsymbol{T}$

(or $\boldsymbol{T}^0$)} as follows:

$$(3)\qquad \begin{cases} x(t,\omega)=y(t,\omega) & \text{if } t<\tau(\omega); \\ x\bigl(\tau(\omega),\omega\bigr)=\lim\limits_{t\downarrow 0} z(t,\omega); & \\ x(t,\omega)=z\bigl(t-\tau(\omega),\omega\bigr) & \text{if } \tau(\omega)\leqq t<\tau(\omega)+\tau'(\omega). \end{cases}$$

This new process is said to be obtained by *hitching* $\{z_t, t<\tau'\}$ *onto* $\{y_t, t<\tau\}$. In view of (7.3) we have then $x\bigl(\tau(\omega),\omega\bigr)=z(0,\omega)$ for a.a. ω if z_0 is defined.

Theorem 1. The process $\{x_t, t<\tau+\tau'\}$ is a M.C. stopped at $\tau+\tau'$ with the minimal state space $\boldsymbol{I}\cup\boldsymbol{I}'$, the substochastic transition matrix whose elements are those of (p_{ij}), $i,j\in\boldsymbol{I}$, and $(p_{i',j'})$, $i',j'\in\boldsymbol{I}'$, and the new ones below:

$$(4)\qquad p_{ii'}(t)=\int_0^t p_{i'}(t-s)\,dL_i(s);\qquad p_{i'i}(t)\equiv 0;\qquad i\in\boldsymbol{I},\ i'\in\boldsymbol{I}'.$$

Proof. The proof is analogous but much simpler than that of Theorem 19.4. We need only calculate

$$\boldsymbol{P}\{x(t_\nu,\omega)=i_\nu,\ 1\leqq\nu\leqq m;\ x(t_\nu,\omega)=i'_\nu,\ m<\nu\leqq n\,|\,x(t_0,\omega)=i_0\}$$

where $0\leqq m<n$ and $0\leqq t_0<\cdots<t_m<\cdots<t_n$. This is equal to

$$\boldsymbol{P}\{y(t_\nu,\omega)=i_\nu,\ 1\leqq\nu\leqq m\,|\,y(t_0,\omega)=i\}\times$$
$$\times\int_{t_m}^{t_{m+1}}\boldsymbol{P}\{z\bigl(t_\nu-\tau(\omega),\omega\bigr)=i'_\nu,\ m<\nu\leqq n\,|\,\tau=s\}\,d_s\boldsymbol{P}\{\tau(\omega)\leqq s\,|\,y(t_m,\omega)=i_m\}$$
$$=\prod_{\nu=0}^{m-1}p_{i_\nu i_{\nu+1}}(t_{\nu+1}-t_\nu)\prod_{\nu=m+1}^{n-1}p_{i'_\nu i'_{\nu+1}}(t_{\nu+1}-t_\nu)\int_{t_m}^{t_{m+1}}p_{i'_{m+1}}(t_{m+1}-s)\,d_sL_{i_m}(s-t_m).$$

The last integral being equal to $p_{i_m i_{m+1}}(t_{m+1}-t_m)$ by (4), the proof is finished. ▯

It is instructive to verify directly that the new matrix is a substochastic transition matrix; this will be left to the reader.

We now proceed to give two specific examples of Theorem 1.

Example 1. Let

$$(5)\qquad 0<q_j<\infty,\qquad \sum_{j=1}^{\infty}\frac{1}{q_j}<\infty;$$

and let $\{\varrho_j, 1\leqq j<\infty\}$ be independent random variables such that the distribution of ϱ_j is e_{q_j} defined in (16.14). Then for each $k\geqq 1$ the sequence $\{\varrho_j, k\leqq j<\infty\}=\{\varrho_k,\varrho_{k+1},\dots\}$ considered in its natural ordering of increasing index j will be called an *ascending escalator*. Let us put $\tau_0 = 0$, $\tau_i = \sum_{j=1}^{i} \varrho_j$; then $\tau_\infty = \lim_{i\to\infty} \tau_i$ is finite a.e. by (5). We define

$$y(t,\omega) = i \quad \text{if } \tau_{i-1}(\omega) \leqq t < \tau_i(\omega).$$

Then $\{y_t, t<\tau\}$ is a stopped M.C.; in fact it is a minimal chain as defined in § 19, whose Q-matrix is given by $q_{i,i+1} = -q_{ii} = q_i$, $i \geqq 1$.

A *descending escalator* is obtained from an ascending one by reversing the ordering of the random variables. We take a new set of indices $\{j', \infty > j \geqq 1\}$ with corresponding $\{q_{j'}\}$ satisfying conditions analogous to (5). Let $\{\varrho_{j'}, \infty > j \geqq 1\} = \{\dots \varrho_{k+1}, \varrho_k, \dots, \varrho_1\}$ be independent random variables such that the distribution function of $\varrho_{j'}$ is $e_{q_{j'}}$. Put $\tau'_i = \sum_{j=i}^{\infty} \varrho_{j'}$ and define

$$z(t,\omega) = i' \quad \text{if } \tau'_{i+1}(\omega) \leqq t < \tau'_i(\omega).$$

Then $\{z_t, 0<t<\tau'_1\}$ is an *open* M.C. stopped at τ'_1. If we hitch this onto $\{y_t, t<\tau_\infty\}$ according to Theorem 1 we obtain a M.C. stopped at $\tau_\infty + \tau'_1$, given as follows:

$$\text{(6)} \qquad x(t,\omega) = \left\{\begin{array}{ll} i & \text{if } \tau_{i-1}(\omega) \leqq t < \tau_i(\omega), \\ \infty & \text{if } t = \tau_\infty(\omega), \\ i' & \text{if } \tau_\infty(\omega) + \tau'_{i+1}(\omega) \leqq t < \tau_\infty(\omega) + \tau'_i(\omega); \end{array}\right\} \qquad i \geqq 1.$$

Observe the appearance of the adjoined state ∞ at the instant of hitching. The substochastic transition matrix of $\{x_t, t<\tau_\infty+\tau'_1\}$ is a solution, but not the minimal one, of the two systems of differential equations in § 17, corresponding to the conservative Q-matrix whose non-vanishing elements are given by

$$q_{i,i+1} = -q_{ii} = q_i, \qquad q_{i',(i-1)'} = -q_{i'i'} = q_{i'}, \qquad i \geqq 1.$$

This follows from Theorem 17.5.

In general, at the end of an escalator, we can hitch on a *new* one (with new states) of either kind, irrespective of the preceding one. This procedure can be repeated transfinitely by suitable choice of the q_i's. Thus, given any ordinal type of denumerable cardinality, we can construct a M.C. in which $S_\infty(\omega)$ for every ω is a set of this type. We leave these possibilities to the reader, but shall treat one case in more detail in Example 2 below. It should be noted that in all these constructions alluded to above, every state appears only once (and so is inessential except the final absorbing state). If however escalators with the same states are repeated then the construction must stop at the first infinite ordinal, since an infinite series of independent and identically distributed non-vanishing random variables diverges with probability one; cf. Theorem 19.4.

Example 2. The simplest stopped M.C. has a single state ϑ with $p_{\vartheta\vartheta}(t) = e^{-q_\vartheta t}$, $0 < q_\vartheta < \infty$. If we hitch a descending escalator with a final absorbing state 0 onto this we obtain a M.C. For the sake of convenient notation we shall make the convention that $\vartheta > j$ for every j, $0 \leqq j < \infty$, let ϱ_ϑ be independent of $\{\varrho_j,\ \infty > j \geqq 1\}$ and have the distribution e_{q_ϑ}, and write $\sum\limits_{j=i}^{\vartheta}$ to denote a sum over all $j \geqq i$ including ϑ. Define

$$
(7)\qquad \begin{cases}
x(t,\omega) = \vartheta & \text{if } 0 \leqq t < \varrho_\vartheta(\omega);\\
x(t,\omega) = \infty & \text{if } t = \varrho_\vartheta(\omega);\\
x(t,\omega) = i & \text{if } \sum\limits_{j=i+1}^{\vartheta} \varrho_j(\omega) \leqq t < \sum\limits_{j=i}^{\vartheta} \varrho_j(\omega), \qquad 1 \leqq i < \vartheta;\\
x(t,\omega) = 0 & \text{if } \sum\limits_{j=1}^{\vartheta} \varrho_j(\omega) \leqq t.
\end{cases}
$$

We leave to the reader a direct verification that this defines a M.C. but proceed in a different manner as follows. Given $\{q_j\}$ satisfying (1) and $q_0 = 0$, $0 < q_\vartheta < \infty$ we shall exhibit a transition matrix (p_{ij}), $0 \leqq i, j \leqq \vartheta$, suggested of course by (7), which satisfies

$$
(8)\qquad p'_{ij}(0) = q_{ij} = \begin{cases}
-q_j & \text{if } i = j\\
q_{j+1} & \text{if } i = j+1,\\
0 & \text{otherwise}.
\end{cases}
$$

(In particular $p'_{\vartheta j}(0) = q_{\vartheta j} = 0$ if $j \neq \vartheta$.) Consider now a M.C. with this (p_{ij}) as transition matrix and the initial distribution $p_\vartheta = 1$, and take the x_+ version of it defined in § 7 for the sake of definiteness. Our previous analysis of sample functions, in particular Theorems 15.2 and 15.6, will show that for this M.C. the sample functions are indeed described by (7). Here the adjoined state ∞ may be eliminated if we take instead of the x_+ version the x_- or $\underline{x}$ version.

Let $D_{\vartheta,\vartheta+1}(t) \equiv 1$, and

$$
D_{ij}(t) \equiv 1 \quad \text{if } i < j \leqq \vartheta; \qquad D_{ij}(t) = \prod_{j \leqq k \leqq i}{}^{*} e_{q_k}(t) \quad \text{if } j \leqq i \leqq \vartheta;
$$

where $\prod^*$ denotes repeated convolution. The desired (p_{ij}) is then given as follows:

$$
(9)\qquad \begin{cases}
p_{ij}(t) \equiv 0, & 1 \leqq i < j \leqq \vartheta;\\
p_{ij}(t) = D_{i,j+1}(t) - D_{ij}(t), & 1 \leqq j \leqq i \leqq \vartheta;\\
p_{i0}(t) = D_{i1}(t),\ p_{0j}(t) \equiv \delta_{0j}, & 0 \leqq i, j \leqq \vartheta.
\end{cases}
$$

To verify that (p_{ij}) is a transition matrix, it is trivial that $p_{ij}(t) \geqq 0$ and $\sum\limits_{j=0}^{\vartheta} p_{ij}(t) \equiv 1$. Furthermore we have

$$
D'_{ij}(t) = \int_0^t q_j[1 - e_{q_j}(t-s)]\, dD_{i,j+1}(s) = q_j[D_{i,j+1}(t) - D_{ij}(t)] = p_{ij}(t)\, q_j
$$

so that it follows from the definition that, if we set $p_{i,\vartheta+1}(t) \equiv 0$ for every $i \leqq \vartheta$ and $q_{\vartheta+1}=0$,

$$p'_{ij}(t)=p_{i,j+1}(t)q_{j+1}-p_{ij}(t)q_j, \quad 0 \leqq i, j \leqq \vartheta; \tag{10}$$

namely that the second system (2) of § 17 holds. Let us put

$$p_{ij}(s,t)=\sum_{k=0}^{\vartheta} p_{ik}(s)p_{kj}(t).$$

We have by (10), since the differentiated series converges uniformly,

$$\begin{aligned}\frac{\partial}{\partial t} p_{ij}(s,t) &= \sum_k p_{ik}(s)p'_{kj}(t) = \sum_k p_{ik}(s)\{p_{k,j+1}(t)q_{j+1}-p_{kj}(t)q_j\} \\ &= p_{i,j+1}(s,t)q_{j+1}-p_{ij}(s,t)q_j.\end{aligned}$$

Thus for each s and i, both $\{p_{ij}(s,t), 0 \leqq j \leqq \vartheta\}$ and $\{p_{ij}(s+t), 0 \leqq j \leqq \vartheta\}$ satisfy the system of equations,

$$z_{ij}(t) \equiv 0, \quad i<j; \qquad z'_{ij}(t)=z_{i,j+1}(t)q_{j+1}-z_{ij}(t)q_j, \quad i \geqq j.$$

Since this system has obviously a unique solution for given $\{z_{ij}(0), 0 \leqq j \leqq \vartheta\}$ it follows that $p_{ij}(s+t) \equiv p_{ij}(s,t)$ identically in s and t. We have thus proved (1.C) and therefore (p_{ij}), $0 \leqq i, j \leqq \vartheta$, defined in (9) is a transition matrix. Since $p_{jj}(t) \equiv e^{-q_j t}$, it is standard and (8) follows at once from (10).

The main feature of this example is the fact that

$$0=\sum_{j \neq \vartheta} q_{\vartheta j}<q_\vartheta;$$

see Theorems 2.6 and 15.6. It is easy to modify this example slightly to make the difference $q_\vartheta-\sum_{j \neq \vartheta} q_{\vartheta j}$ equal to any number between 0 and q_ϑ. We need only, at the exit from ϑ, present the alternative of either entering the absorbing state 0 directly or descending by the escalator.

More interesting ramifications of Example 2 can be obtained in the following way. Escalators will be called *replicas* of each other iff they are mutually independent and corresponding random variables have the same distribution. Choose an arbitrary i, $0 \neq i \neq \vartheta$. Instead of entering $i-1$ at the exit from i we have the alternative of either (with probability λ_i) entering $i-1$ or (with probability $1-\lambda_i$) running through a replica of the descending escalator $\{\varrho_j, \infty > j \geqq 1\}$ just as previously at the exit from ϑ. The same alternative must then be present at each exit from i to preserve the Markovian character of the process. The new process can be easily analysed by an obvious decomposition. A set of sample functions of probability λ_i is defined as in (7); and for each $n \geqq 1$ a set of sample functions of probability $\lambda_i(1-\lambda_i)^n$ has exactly n replicas of the descending escalator $\{\varrho_j, \infty > j \geqq i\}$ succeeding each other at the first exit from i. A similar analysis applies if the same kind of ramification is made at a finite number of states. The analysis becomes more involved if this is done at an infinite number of states; we may then as well suppose that it is done at every i since we may put $\lambda_i = 1$ for certain i. Without going into details let us indicate an analytic method. Define a sequence of distribution functions $\{F_{\infty j}, 0 < j < \vartheta\}$ as follows:

$$F_{\infty j} = F_{\infty, j+1} * e_{q_{j+1}} * [\lambda_{j+1} \epsilon + (1 - \lambda_{j+1}) F_{\infty j}]$$

or equivalently

$$F_{\infty j} = \sum_{n=0}^{\infty} [F_{\infty, j+1} * e_{q_{j+1}}]^{(n+1)*} \lambda_{j+1} (1 - \lambda_{j+1})^n .$$

A sufficient condition that $F_{\infty j} \not\equiv 0$ for all j and that $\lim\limits_{j \to \infty} F_{\infty j} = \epsilon$ is

$$\sum_j (\lambda_1 \ldots \lambda_j q_j)^{-1} < \infty .$$

Next we define

$$\left.\begin{aligned} &F_{ij} = e_{q_i} * [\lambda_i F_{i-1, j} + (1 - \lambda_i) F_{\infty j}]; \\ &F_{\vartheta j} = e_{q_\vartheta} * F_{\infty j}; \qquad F_{i\vartheta} \equiv 0; \end{aligned}\right\} \qquad 0 < i, j < \vartheta .$$

It is surmised that these distribution functions are the first entrance time distributions of the M.C., from which the matrix (p_{ij}) is to be calculated by (13.3). It is to be verified that the latter is indeed a transition matrix with the required properties.

Example 3. In this example the state space is the set $\{j, 0 \leqq j < \infty\}$ and there is exactly one instantaneous state 0. We attempt to construct a transition matrix (p_{ij}) satisfying

$$\text{(11)} \qquad \left\{\begin{aligned} &-p'_{00}(0) = q_0 = \infty, && p'_{0j}(0) = q_{0j} = 1, \\ &-p'_{jj}(0) = q_j, && p'_{j0}(0) = q_{j0} = q_j, \end{aligned}\right\} \qquad 1 \leqq j < \infty;$$

where $0 < q_j < \infty$ and

$$\text{(12)} \qquad \sum_{j=1}^{\infty} \frac{1}{q_j} < 1 .$$

The condition (12) will be discussed shortly. It follows from (11) that $p'_{ij}(0) = q_{ij} = 0$ if $i, j \geqq 1$ and $i \neq j$. Thus a stable state cannot be entered by a jump from another state. It is still possible to enter a stable state by a pseudo-jump from ∞, but we seek to rule this out by requiring that any stable state can be entered only by a pseudo-jump from the instantaneous state. By Theorem 16.3, this requirement amounts to setting

$$\text{(13)} \qquad p_{ij}(t) = \int_0^t p_{i0}(s) q_{0j} e^{-q_j (t-s)} \, ds + \delta_{ij} e^{-q_j t}, \qquad i \geqq 0, j \geqq 1 .$$

Hence (2_{ij}) of § 17 holds if $i \geqq 0, j \geqq 1$. It follows from (11) and Theorem 17.2 that (1_{ij}) of § 17 holds if $i \geqq 1$, $j \geqq 0$. Thus both (1_{ii}) and (2_{ii}) hold if $i \geqq 1$ and upon comparison we have

$$p_{i0}(t) = q_i\, p_{0i}(t), \quad i \geqq 1. \tag{14}$$

Equations (13) and (14) determine all p_{ij} as soon as p_{00} is given, and the latter must satisfy the equation $\sum_{j=0}^{\infty} p_{0j}(t) \equiv 1$, namely

$$p_{00}(t) + \int_0^t p_{00}(s)\, h(t-s)\, ds \equiv 1 \tag{15}$$

where

$$h(t) = \sum_{j=1}^{\infty} e^{-q_j t}.$$

Taking Laplace transforms in (15) and using the uniqueness theorem for the Laplace transform we see that equation (15) has a unique solution $p_{00}(\cdot)$ which is clearly continuous. Under the assumption (12) it is easy to show that $p_{00}(t) \geqq 0$ for all t; for otherwise let t_0 be the infimum of t such that $p_{00}(t) < 0$, then $p_{00}(t) \leqq 1$ for $0 \leqq t \leqq t_0$ and equation (15) with $t = t_0$ is impossible. The assertion is actually true even if (12) is relaxed to read $\sum_{j=1}^{\infty} q_j^{-1} < \infty$; see REUTER [1] where a more general result is proved. For our purpose the condition (12) will suffice.

To recapitulate, we define (p_{ij}) by (15), (13) and (14) in this order and proceed to show that it is a transition matrix satisfying (11). We know already that $p_{ij}(t) \geqq 0$ and $\sum_{j=0}^{\infty} p_{0j}(t) = 1$. Let us write $[f \circ g](t) = \int_0^t f(t-s)\, g(s)\, ds$; we have by the definitions, if $i \geqq 1$,

$$\begin{aligned} \sum_{j=0}^{\infty} p_{ij}(t) &= p_{i0}(t) + [p_{i0} \circ h](t) + e^{-q_i t} \\ &= [p_{00} \circ e'_{q_i}](t) + [p_{00} \circ h \circ e'_{q_i}](t) + e^{-q_i t} \\ &= [1 \circ e'_{q_i}](t) + e^{-q_i t} = 1. \end{aligned}$$

Next, as already noted above, (2_{ij}) holds, namely:

$$p'_{ij}(t) = p_{i0}(t)\, q_{0j} - p_{ij}(t)\, q_j, \quad i \geqq 0, j \geqq 1. \tag{16}$$

These equations determine $\{p_{ij}, i \geqq 0, j \geqq 1\}$ for given $\{p_{i0}, i \geqq 0\}$. Hence an argument similar to that in Example 2 shows that

$$p_{ij}(s+t) = \sum_k p_{ik}(s)\, p_{kj}(t) \quad i \geqq 0, j \geqq 1.$$

The above equation then holds also if $i=j=0$, since $p_{00}(s+t) = 1-\sum_{j=1}^{\infty} p_{0j}(s+t)$. Finally we have clearly $\lim_{t\to 0} p_{ij}(t)=\delta_{ij}$; and

$$q_0 = \lim_{t\to 0} \frac{1-p_{00}(t)}{t} = \lim_{t\to 0} \frac{1}{t}\int_0^t p_{00}(s)\, h(t-s)\, ds = \lim_{t\to 0} h(t) = \infty .$$

The rest of (11) follows from (16) and (14).

We have therefore constructed a transition matrix in which exactly one state is instantaneous. For any M.C. with this transition matrix our previous analysis of sample functions yields the following description for a.a. ω. The interval $\boldsymbol{T}$ is composed of the stable intervals which are everywhere dense and the set $S_0(\omega)$ which is dense-in-itself, non-dense and of positive measure. No adjoined state ∞ is needed if one of the three versions $\underline{x}$, x_+ or x_- is used. Each endpoint of a stable interval is a limit point of $S_0(\omega)$, hence also of stable intervals with indices going to infinity as the endpoint is approached from outside the interval. If in particular the $\underline{x}$ version is taken then (because of the labelling of the instantaneous state with the smallest index) all the endpoints of stable intervals belong to $S_0(\omega)$. The set $S_0(\omega)$ is then simply the closure of the set of these endpoints; it is thus a perfect set and gives a perfect example of a (generalized) Cantor set.

It remains to analyse the ordering of the stable intervals. For this purpose we shall calculate the ${}_HF_{ij}(\infty)$ and ${}_HP_{ij}(\infty)$ defined in § 11. Let us first note that

$$\int_0^{\infty} p_{00}(t)\, dt = \infty$$

from (15) so that by Theorem 10.4 the states form a recurrent class, and all $F_{ij}(\infty)=1$. It is obvious, since 0 is entered necessarily at the exit from i, that ${}_HF_{ij}(\infty) = {}_HF_{0j}(\infty)$ for any i, j and H. Let H be a finite set of n_H states, and let j and k be distinct stable states not in H. We have from (11.20)

$${}_HF_{0j}(\infty) = {}_{k,H}F_{0j}(\infty) + {}_HF_{kj}(\infty)\, {}_{j,H}F_{0k}(\infty) .$$

Interchanging j and k in this equation, and using induction on the number of states in H, we obtain

$$(17) \qquad {}_HF_{ij}(\infty) = \frac{1}{n_H+1},$$

for every i, j and finite H not containing 0. Henceforth H shall always exclude 0. Thus ${}_HF_{ij}(\infty)=0$ if H is an infinite set, which means that with probability one any given infinite set of states is entered between the exit from a given state and the entrance into a given state. To calculate ${}_HP_{ij}(\infty)$ we have first from (13), and using the notation of § 13,

$$P_{jj}(t) = \int_0^t p_{j0}(s)\, q_j^{-1}[1-e^{-q_j(t-s)}]\, ds + q_j^{-1}[1-e^{-q_j t}], \quad j \neq 0\,;$$

from which and (13.5) it follows that

$$\lim_{t\to\infty} \frac{P_{jj}(t)}{P_{00}(t)} = \lim_{t\to\infty} \frac{P_{jj}(t)}{P_{j0}(t)} = \frac{1}{q_j}. \tag{18}$$

Owing to the special nature of the process, we have ${}_0P_{jj}(\infty) = q_j^{-1}$. Hence we obtain from (18) and (13.12)

$${}_jP_{00}(\infty) = 1, \quad j \neq 0. \tag{19}$$

Next we have from (11.19) if $k \notin H$,

$${}_HP_{00}(\infty) = {}_HP_{k0}(\infty)\,{}_HF_{0k}(\infty) + {}_{k,H}P_{00}(\infty). \tag{20}$$

Since ${}_HP_{k0}(\infty) = {}_HP_{00}(\infty)$ by the special nature of the process we obtain from (20) and (17)

$${}_{k,H}P_{00}(\infty) = {}_HP_{00}(\infty)\, \frac{n_H}{n_H+1}.$$

We may thus again use induction on the number of states in H and conclude from (19) that

$${}_HP_{00}(\infty) = \frac{1}{n_H}. \tag{21}$$

From now on we shall suppose $\mathbf{P}\{x_0(\omega)=0\}=1$. According to Theorem 14.3 and (19) above, the random variables $\{{}_jC_{00}, j \geq 1\}$ defined before Theorem 14.3 have the common distribution function e_1. We prove that they are independent. Let $\{t_j, 1 \leq j \leq k\}$ be arbitrary positive numbers; without loss of generality we may suppose that $t_1 < \cdots < t_k$. Using the notation introduced before Theorem 14.3 and applying the strong Markov property to $\gamma_0(t_1), \ldots, \gamma_0(t_k)$ successively, we see that

$$\mathbf{P}\{{}_jC_{00}(\omega) > t_j,\ 1 \leq j \leq k\} = \prod_{j=1}^{k} \mathbf{P}({}_{H_j}C_{00}(\omega) > t_j - t_{j-1}\}$$

where $t_0 = 0$ and $H_j = \{j, \ldots, k\}$. By Theorem 14.3, the above probability is equal to

$$\prod_{j=1}^{k} e^{-j(t_j - t_{j-1})} = \prod_{j=1}^{k} e^{-t_j} = \prod_{j=1}^{k} \mathbf{P}\{{}_jC_{00}(\omega) > t_j\}.$$

This being true for every k, the asserted independence is proved.

It follows from the result above that the time spent in the state 0 until the first entrance into an interval with a label from 1 to k has the distribution e_k; and that for each j from 1 to k this interval is a j-interval with probability k^{-1}. Applying the same result again to the terminal point $\alpha(j)$ of this interval, since $x_+(\alpha(j;\omega),\omega)=0$, we see that the time spent in the state 0 until the entrance into the next interval with a distinct label from 1 to k has the distribution e_{k-1}; and for each such label j' this second interval is a j'-interval with probability $(k-1)^{-1}$. The argument can now be repeated at the terminal point of the second interval, and so on. These considerations lead to the following construction. Define a M.C. $\{\hat{x}_t, t\geqq 0\}$ with $k+1$ states $\{0, 1, \ldots, k\}$ whose Q-matrix is given by

$$\hat{q}_0=k, \hat{q}_{0j}=1, \hat{q}_i=k-1, \hat{q}_{ij}=1, \hat{q}_{i0}=0, \qquad 1\leqq i\neq j\leqq k$$

and whose initial distribution is concentrated at 0. Let $\{\tau_n(j); n\geqq 1, j\geqq 1\}$ and $\{\hat{\tau}_n(j); n\geqq 1, 1\leqq j\leqq k\}$ be the entrance times for the $\{x_t\}$ and $\{\hat{x}_t\}$ processes respectively, as defined in § 15. Let $\hat{\varrho}_n(j)=\hat{\tau}_{k+1}(j)-\hat{\tau}_n(j)$ be the n-th return time of the state j in $\{\hat{x}_t\}$. Furthermore let

$$\mu_n(j;\omega)=\mu\left[S_0(\omega)\cap\left(\tau_n(j;\omega), \tau_{n+1}(j;\omega)\right)\right]$$

so that $\mu_n(j)$ is the time spent in the state 0 between the n-th and $n+1$-st entrance times of the state j. The succession of the states $\{1, \ldots, k\}$ in the M.C. $\{x_t\}$ is the same as in $\{\hat{x}_t\}$, and the *time spent in the state* 0 between a change of states in $\{x_t\}$ is the same as the ordinary time between the corresponding change of states in $\{\hat{x}_t\}$. It follows that the mapping $\mu_n(j)\Leftrightarrow\hat{\varrho}_n(j)$, $n\geqq 1$, $1\leqq j\leqq k$, is probability-preserving.

Now the M.C. $\{\hat{x}_t\}$ can be realized in the following way. For each j, $1\leqq j\leqq k$, let $\{\zeta_n(j), n\geqq 1\}$ be independent random variables with the common distribution e_1 and let $\sigma_n(j)=\sum_{\nu=1}^{n}\zeta_\nu(j)$ $(\sigma_0(j)=0)$. Define $\{x_t^{(j)}, t\geqq 0\}$ as follows: $x^{(j)}(t,\omega)=n$ if $\sigma_n(j;\omega)\leqq t<\sigma_{n+1}(j;\omega)$, $0\leqq n<\infty$. Thus $\{x_t^{(j)}\}$ is a Poisson process with parameter value equal to one. For each ω, $\sigma_n(j;\omega)$ is the n-th point of jump in the j-th Poisson process. Let the Poisson processes $\{x_t^{(j)}\}$, $1\leqq j\leqq k$, be independent of each other and arrange the set $\{\sigma_n(j;\omega); n\geqq 1, 1\leqq j\leqq k\}$ in an increasing sequence $\{\eta_n(\omega), n\geqq 1\}$. If N is the ω-set where $\sigma_n(j;\omega)=\sigma_{n'}(j';\omega)$ for some $(j,n)\neq(j',n')$, then N is a null set. Define a new process $\{\breve{x}_t, t\in\boldsymbol{T}\}$ as follows: $\breve{x}(t,\omega)=0$ if $0\leqq t<\eta_1(\omega)$; $\breve{x}(t,\omega)=j$ if $\eta_n(\omega)\leqq t<\eta_{n+1}(\omega)$ and if $\eta_n(\omega)=\sigma_m(j;\omega)$ for a certain m. Thus $\breve{x}(t,\omega)$ is uniquely defined if $\omega\in\mathsf{N}$. It is easy to see that $\{\breve{x}_t\}$ is a M.C. with the same states, Q-matrix and initial distribution as $\{\hat{x}_t\}$. Hence by Theorem 18.3 and Corollary 2 to Theorem 19.1 the two processes have the same finite-dimensional joint distributions. In other words, the mapping $\hat{x}_t\Leftrightarrow\breve{x}_t$, $t\in\boldsymbol{T}$, is probability-preserving. In particular the mapping $\hat{\varrho}_n(j)\Leftrightarrow\breve{\varrho}_n(j)$, $n\geqq 1$, $1\leqq j\leqq k$, is probability-preserving, where $\breve{\varrho}_n(j)$ is the n-th return time of the state j in $\{\breve{x}_t\}$. But by construction we have $\breve{\varrho}_n(j)=\zeta_n(j)$ so that the random variables $\{\breve{\varrho}_n(j); n\geqq 1, 1\leqq j\leqq k\}$ are mutually independent.

Hence the random variables $\{\hat{\varrho}_n(j),\ n\geqq 1,\ 1\leqq j\leqq k\}$ and consequently also $\{\mu_n(j);\ n\geqq 1,\ 1\leqq j\leqq k\}$ are mutually independent. This being true for every k we have proved that the random variables $\{\mu_n(j);\ n\geqq 1, j\geqq 1\}$ are mutually independent and have the common distribution e_1.

The ordering of the stable intervals in the original process $\{x_t\}$ can now be pictured in the following way. Consider a sequence of mutually independent Poisson processes (with parameter value one) to each of which a different color is assigned. Each sample function of the j-th Poisson process may be visualized as the half-line $[0, \infty)$ marked with a sequence of marks $\{1, 2, \ldots\}$ in the j-th color, the successive marks being placed at the successive points of jump of the sample function. If we juxtapose the denumerable number of marked half-lines corresponding to the same ω, the merged line has a denumerable and everywhere dense set of marks in the infinite variety of colors. Each such ordering of marks represents an ordering of the stable intervals in the $\{x_t\}$ process, and the distance between two given marks represents the time spent in the instantaneous state between the corresponding intervals. To obtain the sample functions $x(\cdot, \omega)$ we must replace each mark by an interval of the same color, the lengths of all the intervals being independently distributed and each interval of the j-th color having a length distributed according to e_{q_j}.

The mathematical construction answering the above description will now be given. Let $\{\mu_n(j), \lambda_n(j);\ n\geqq 1,\ j\geqq 1\}$ be mutually independent random variables with the common distribution e_1. Given j and n, define $m=m(j, n, k; \omega)$ for every k by

$$\sum_{\nu=1}^{m} \mu_\nu(k; \omega) < \sum_{\nu=1}^{n} \mu_\nu(j; \omega) \leqq \sum_{\nu=1}^{m+1} \mu_\nu(k; \omega)$$

and let

$$\tau_n(j; \omega) = \sum_{\nu=1}^{n} \mu_\nu(j; \omega) + \sum_{k=1}^{\infty} \sum_{\nu=1}^{m(j,n,k;\omega)} \lambda_\nu(k; \omega).$$

Define $x(0, \omega)=0$; and for each $j\geqq 1$ define

$$x(t, \omega)=j \quad \text{if} \quad \tau_n(j, \omega)\leqq t<\tau_n(j, \omega)+\lambda_n(j, \omega);$$

finally define $x(t, \omega)=0$ for all other values of t. We leave it to the reader to verify that this definition is consistent for a.a. ω and that the sample functions so constructed are precisely those prescribed by our preceding analysis.

Example 4. This example is a kind of superposition of Examples 2 and 3. It is obtained from Example 3 by replacing each stable interval by a descending escalator. The state space is $\{0, j_n; j\geqq 1,\ n\geqq 1\}$. Let

$0 < q_{j_n} < \infty$ and

$$\sum_{j_n \neq 0} \frac{1}{q_{j_n}} < 1. \tag{22}$$

This condition can be relaxed in the same way as (12). Let $\{\varrho_{j_n};\, j \geqq 1, n \geqq 1\}$ be independent random variables with the distribution functions $e_{q(j_n)}$ where $q(j_n) = q_{j_n}$. Let

$$\left|\begin{aligned} &F^j_{mn}(t) = \begin{cases} \boldsymbol{P}\left\{\sum\limits_{\nu=n}^{m} \varrho_{j_\nu}(\omega) \leqq t\right\}, & 1 \leqq n \leqq m \leqq \infty; \\ 1, & 1 \leqq m < n \leqq \infty; \end{cases} \\ &r^j_{mn}(t) = F^j_{m,\,n+1}(t) - F^j_{mn}(t) = \boldsymbol{P}\left\{\sum_{\nu=n+1}^{m} \varrho_{j_\nu}(\omega) \leqq t < \sum_{\nu=n}^{m} \varrho_{j_\nu}(\omega)\right\}; \end{aligned}\right. \tag{23}$$

where the symbols with $m = \infty$ are the limits as $m \to \infty$, which exist on account of (22). Let us observe that

$$\sum_{n=1}^{m} r^j_{mn}(t) = 1 - F^j_{m1}(t), \quad 1 \leqq m \leqq \infty.$$

We define p_{00} as the unique solution of the following equation:

$$p_{00}(t) + \int_0^t p_{00}(s) \sum_{j=1}^{\infty} \sum_{n=1}^{\infty} r^j_{\infty n}(t-s)\, ds \equiv 1; \tag{24}$$

furthermore we define

$$\left|\begin{aligned} &p_{i_m 0}(t) = \int_0^t p_{00}(t-s)\, d F^i_{m1}(s), && i_m \neq 0; \\ &p_{0 j_n}(t) = \int_0^t p_{00}(t-s)\, r^j_{\infty n}(s)\, ds, && j_n \neq 0; \\ &p_{i_m j_n}(t) = \int_0^t p_{i_m 0}(t-s)\, r^j_{\infty n}(s)\, ds + \delta_{ij}\, r^j_{mn}(t), && i_m \neq 0 \neq j_n. \end{aligned}\right. \tag{25}$$

We leave it to the reader to verify, in the same manner as in Example 3, that (23)–(25) define a transition matrix satisfying

$$-p'_{00}(0) = \infty, \quad p'_{0 j_n}(0) = 0,$$
$$-p'_{j_n j_n}(0) = p'_{j_n j_{n-1}}(0) = q_{j_n}, \quad j_n \neq 0;$$

where $j_0 = 0$ for each $j \geqq 1$. The feature which distinguishes this example from the preceding one is that here we have

$$\sum_{j_n \neq 0} p'_{0 j_n}(0) = 0 < \infty = -p'_{00}(0).$$

Example 5. In both Examples 3 and 4 there is exactly one instantaneous state. It is now an easy matter to derive examples in which the number of instantaneous states is arbitrarily given. The following method of constructing a *product M.C.* is general. Let $\{x_t^*, t \in \boldsymbol{T}\}$ and $\{x_t^{**}, t \in \boldsymbol{T}\}$ be two mutually independent M.C.'s with state spaces $\boldsymbol{I}^*$ and $\boldsymbol{I}^{**}$ and transition matrices $(p_{i^*j^*})$ and $(p_{i^{**}j^{**}})$ respectively. Define a new process $\{x_t, t \in \boldsymbol{T}\}$ with state space $\boldsymbol{I}$ consisting of the ordered pairs (i^*, i^{**}) where $i^* \in \boldsymbol{I}^*$ and $i^{**} \in \boldsymbol{I}^{**}$, defined as follows:

$$x(t, \omega) = (i^*, i^{**}) \quad \text{if} \quad x^*(t, \omega) = i^* \quad \text{and} \quad x^{**}(t, \omega) = i^{**}.$$

(All pairs of the form (i^*, ∞^{**}) or (∞^*, i^{**}) or (∞^*, ∞^{**}) where ∞^* and ∞^{**} are the adjoined states, if any, of $\{x_t^*\}$ and $\{x_t^{**}\}$ respectively, may be merged into one adjoined state of $\{x_t\}$.) Clearly $\{x_t, t \in \boldsymbol{I}\}$ is a M.C. whose transition matrix (p_{ij}) is given as follows: If $i = (i^*, i^{**})$ and $j = (j^*, j^{**})$, then

$$p_{ij}(t) \equiv p_{i^*j^*}(t)\, p_{i^{**}j^{**}}(t).$$

This new transition matrix will be called the *direct product* of the two given ones.

Now let $\{x_t^*, t \in \boldsymbol{T}\}$ have exactly one instantaneous state as in Example 3 or 4; and $\{x_t^{**}, t \in \boldsymbol{T}\}$ have exactly N states (all stable of course). Since

$$p'_{ij}(0) = p'_{i^*j^*}(0)\, p_{i^{**}j^{**}}(0) + p_{i^*j^*}(0)\, p'_{i^{**}j^{**}}(0),$$

it follows that exactly N states of $\{x_t, t \in \boldsymbol{T}\}$ are instantaneous, namely those (i^*, i^{**}) where i^* is instantaneous and i^{**} is arbitrary.

Example 6. In this example all the states are instantaneous. Let $\{x_t^{(n)}, t \in \boldsymbol{T}\}$, $n \geqq 1$, be a sequence of mutually independent M.C.'s. Each M.C. has the state space consisting of two states $\{0, 1\}$ and the transition matrix $(p_{ij}^{(n)})$, $i = 0, 1$, is determined by the numbers $q_0^{(n)}$ and $q_1^{(n)}$ in our customary notation .In fact, we have, by solving either system (1) or (2) of § 17,

$$p_{00}^{(n)}(t) = \frac{q_1^{(n)}}{q_0^{(n)} + q_1^{(n)}} + \frac{q_0^{(n)}}{q_0^{(n)} + q_1^{(n)}}\, e^{-(q_0^{(n)} + q_1^{(n)})t}. \tag{26}$$

We suppose that $q_0^{(n)}$ and $q_1^{(n)}$ are positive for each n, and

$$\sum_n \frac{q_0^{(n)}}{q_0^{(n)} + q_1^{(n)}} < \infty; \tag{27}$$

$$\sum_n q_0^{(n)} = \infty. \tag{28}$$

Consider now the *infinite product process* $\{X_t, t \in \boldsymbol{T}\}$ where $X_t = \{x_t^{(n)}, 1 \leqq n < \infty\}$, whose state space is contained in the set of infinite sequences $\boldsymbol{\epsilon} = \{\varepsilon_n, 1 \leqq n < \infty\}$ of 0's and 1's. The corresponding transition matrix is given by

$$p_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}(t) = \boldsymbol{P}\{X_{s+t}(\omega) = \boldsymbol{\epsilon}' \mid X_s(\omega) = \boldsymbol{\epsilon}\} = \prod_{n=1}^{\infty} p_{\varepsilon_n \varepsilon'_n}^{(n)}(t). \tag{29}$$

Let $\boldsymbol{P}\{x_0^{(n)}(\omega)=0\}=1$ for every n so that $\boldsymbol{P}\{X_0(\omega)=\boldsymbol{0}\}=1$ where $\boldsymbol{0}$ is the sequence of all 0's. We have from (26) and (27)

$$\sum_n p_{01}^{(n)}(t) \leqq \sum_n \frac{q_0^{(n)}}{q_0^{(n)}+q_1^{(n)}} < \infty.$$

Hence by the Borel-Cantelli lemma, for each fixed $t>0$ and a.a. ω, only finitely many $x_t^{(n)}$'s are equal to 1. Thus if $\boldsymbol{\epsilon}$ contains an infinite number of 1's then $\boldsymbol{P}\{X_t(\omega)=\boldsymbol{\epsilon}\,|\,X_0(\omega)=\boldsymbol{0}\}=0$. In other words, each such $\boldsymbol{\epsilon}$ is a fictitious state for the $\{X_t\}$ process in the definition of § 4. On the other hand, if $\boldsymbol{\epsilon}$ contains only a finite number of 1's then $\boldsymbol{P}\{X_t(\omega)=\boldsymbol{\epsilon}\,|\,X_0(\omega)=\boldsymbol{0}\}>0$ for each $t>0$. Let the denumerable set of infinite sequences of 0's and 1's containing only a finite number of 1's be the minimal state space $\boldsymbol{I}$. (If we regard each such sequence as the dyadic expansion of a real number between 0 and 1, the set $\boldsymbol{I}$ is not a discrete set. The natural compactification on the real line requires the adjoining of a nondenumerable set of non-isolated fictitious states in the sense of § 4. In this representation we have $X_t=\sum_{n=1}^{\infty} x_t^{(n)}2^{-n}$.)

If $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}'$ both belong to $\boldsymbol{I}$, then there exists an $N<\infty$ such that $\varepsilon_n=\varepsilon_n'=0$ for all $n\geqq N$. Hence if $n\geqq N$, $p_{\varepsilon_n\varepsilon_n'}^{(n)}(t)\geqq q_1^{(n)}(q_0^{(n)}+q_1^{(n)})^{-1}$ by (26) and the infinite product in (29) converges by (27) (to a nonzero limit). It follows that we have $\lim_{t\to 0} p_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}(t)=1$ for each $\boldsymbol{\epsilon}$ in $\boldsymbol{I}$ and that $(p_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'})$ is a standard transition matrix. The analytical verification of (1.B) and (1.C) is left to the reader. Thus there is a version of $\{X_t, t\in\boldsymbol{T}\}$ which is well-separable and measurable by Theorem 4.3. For such a version we have for every $s\geqq 0$, $h>0$,

$$\begin{aligned}\boldsymbol{P}\{X(t,\omega)&=\boldsymbol{\epsilon},\ s\leqq t\leqq s+h\,|\,X(s,\omega)=\boldsymbol{\epsilon}\}\\ &=\prod_{n=1}^{\infty}\boldsymbol{P}\{x^{(n)}(t,\omega)=\varepsilon_n,\ s\leqq t\leqq s+h\,|\,x^{(n)}(s,\omega)=\varepsilon_n\}\\ &=\prod_{n=1}^{\infty}e^{-q_{\varepsilon_n}^{(n)}h}=e^{-\sum_{n=1}^{\infty}q_{\varepsilon_n}^{(n)}h}.\end{aligned}$$

Since $\varepsilon_n=0$ for all sufficiently large n, the last term is equal to zero by (28). Hence every state $\boldsymbol{\epsilon}$ is instantaneous.

If all states are instantaneous, then we have for a.e. ω:

$$\overline{S_\infty(\omega)}=\boldsymbol{T}. \tag{30}$$

To prove this we need the following extension of BAIRE's lemma which can be proved in the same way as its usual form (see e.g. KELLEY [1; p. 200]).

Lemma. Let $\{U_n, n \geqq 1\}$ be a sequence of subsets of $\boldsymbol{T}$ such that each U_n is dense in $\boldsymbol{I}$ and has the following property: for any open interval V in $\boldsymbol{T}$, whenever $U_n \cap V$ is not empty it contains a nonempty open interval. Then $\bigcap_{n=1}^{\infty} U_n$ is dense in $\boldsymbol{T}$.

To apply the lemma to the proof of (30), we may suppose that $\boldsymbol{I}$ is the set of positive integers. For each ω and $n \geqq 1$ put $U_n = \bigcup_{n \leqq m \leqq \infty} S_m(\omega)$. If all states are instantaneous, it follows from Theorem 7.4 (β) and the fact that $\mu[S_\infty(\omega)] = 0$ for a.e. ω (Theorem 5.1) that for a.e. ω the sequence $\{U_n(\omega), n \geqq 1\}$ satisfies the conditions of the Lemma. Hence $S_\infty(\omega) = \bigcap_{n=1}^{\infty} U_n(\omega)$ is dense in $\boldsymbol{T}$, as was to be shown. □

Example 7. Even if all states are stable, it is possible that (almost) no sample function has any jumps so that all its discontinuities are of the second kind. Each such function is a singular nondecreasing function in $\boldsymbol{T}$; its graph is obtained by inverting that of a purely discontinuous strictly increasing function whose set of discontinuity in that of all the rationals in $\boldsymbol{T}$, the segments of the jumps being added to the graph. Thus as t increases, $x(t, \omega)$ assumes all values in $\boldsymbol{T}$, each rational being assumed in a nondegenerate interval while the set of irrationals being assumed in a set of Lebesgue measure zero.

To construct such a M.C., let R be the set of rationals in $\boldsymbol{T}$ and let r denote an element of R below. Assign to each r a positive number q_r such that $q_r = q_{r+1}$ for every r and $\sum_{0 \leqq r \leqq 1} q_r^{-1} < \infty$. Thus $\sum_{r \in R} q_r^{-1} = \infty$. Let $\{\tau_r, r \in R\}$ be a countable family of independent random variables such that τ_r has the distribution function e_{q_r}, so that $E\{\tau_r\} = q_r^{-1}$. Put

$$\sigma(s, \omega) = \sum_{r \leqq s} \tau_r(\omega)$$

for $s \in \boldsymbol{T}$. There exists a set Ω_0 with $\boldsymbol{P}(\Omega_0) = 1$ such that if $\omega \in \Omega_0$, then the following statements are all true. For each r, $0 < \tau_r(\omega) < \infty$; for each $s \in \boldsymbol{T}$, $\sigma(s, \omega) < \infty$; $\sigma(\cdot, \omega)$ is strictly increasing in $\boldsymbol{T}$; $\lim_{s \to \infty} \sigma(s, \omega) = \infty$; and

$$(31) \qquad \sigma(s+0, \omega) - \sigma(s-0, \omega) = \begin{cases} \tau_r(\omega) > 0, & \text{if } s = r \in R; \\ 0, & \text{if } s \in \boldsymbol{T} - R. \end{cases}$$

Thus $\sigma(\cdot, \omega)$ has jumps at each r in R and is continuous elsewhere. These properties follow from the proposition that the sum of a countable family of independent positive random variables is finite for a.e. ω if and only if the sum of their expectations is finite; and that if a countable family of positive numbers have a finite sum then any sequence from the family has a convergent sum.

Now put for $\omega \in \Omega_0$ and $t \in \boldsymbol{T}$:

$$x(t, \omega) = s \quad \text{for} \quad t \in [\sigma(s-0, \omega), \sigma(s+0, \omega)], \qquad s \in \boldsymbol{T}. \tag{32}$$

The sample functions of the process $\{x_t, t \in \boldsymbol{T}\}$ have the properties described above. If $s \in \boldsymbol{T} - R$ and $t \in \boldsymbol{T}$, then

$$\boldsymbol{P}\{x(t, \omega) = s\} = \boldsymbol{P}\{\sigma(s, \omega) = t\} = 0$$

by (31) and the (absolute) continuity of the distribution of $\sigma(s, \cdot)$. Next if $0 = t_0 < t_1 < \cdots < t_l$ and $0 = r_0 < r_1 < \cdots < r_l$ in R, we have

$$\begin{aligned} &\boldsymbol{P}\{x(t_\nu, \omega) = r_\nu, 1 \leqq \nu \leqq l\} \\ &\quad = \boldsymbol{P}\Big\{\sum_{r<r_\nu} \tau_r(\omega) \leqq t_\nu \leqq \sum_{r \leqq r_\nu} \tau_r(\omega), 1 \leqq \nu \leqq l\Big\} \\ &\quad = \boldsymbol{P}\Big\{\sum_{r_{\nu-1} \leqq r < r_\nu} \tau_r(\omega) \leqq t_\nu - t_{\nu-1} \leqq \sum_{r_{\nu-1} \leqq r \leqq r_\nu} \tau_r(\omega), 1 \leqq \nu < l\Big\}. \end{aligned} \tag{33}$$

The first equation is a consequence of the definition (32); the second may be proved by induction on l using the following lemma.

Lemma. Let σ, τ and σ' be independent random variables where σ and σ' are positive while τ has the distribution function e_q. Then we have for $0 \leqq t \leqq t'$:

$$\begin{aligned} &\boldsymbol{P}\{\sigma(\omega) \leqq t \leqq \sigma(\omega) + \tau(\omega); \sigma(\omega) + \tau(\omega) + \sigma'(\omega) \leqq t'\} \\ &\quad = \boldsymbol{P}\{\sigma(\omega) \leqq t \leqq \sigma(\omega) + \tau(\omega)\} \boldsymbol{P}\{\tau(\omega) + \sigma'(\omega) \leqq t' - t\}. \end{aligned}$$

To prove the lemma, let F and G be the distribution functions of σ and σ'. Writing as usual $\int_0^t$ for $\int_{[0,t]}$ etc., we see that the left member of (33) is equal to

$$\begin{aligned} &\int_0^t dF(u) \int_{t-u}^{t'-u} q\, e^{-qv}\, dv\, G(t'-u-v) \\ &\quad = \int_0^t dF(u)\, e^{-q(t-u)} \int_0^{t'-t} q\, e^{-qv}\, G(t'-t-v)\, dv \end{aligned}$$

which is equal to the right member of (33). Let us remark that the calculation above is based on the fundamental "memoryless" property of a random variable τ with distribution function e_q, which may be expressed as follows, for $0 \leqq s \leqq t$:

$$\boldsymbol{P}\{\tau \leqq t \mid \tau > s\} = \boldsymbol{P}\{\tau \leqq t - s\};$$

or analytically as in the above:

$$e^{qs} \int_s^t e^{-qv} G(v)\, dv = \int_0^{t-s} e^{-qv}\, G(v+s)\, dv.$$

This property of e_q, as the sojourn time distribution of a stable state, is implicit in Theorem 15.2 and its consequences. It may be used to give direct analytical verifications that the escalators in Examples 1 and 2, and the process constructed in Theorem 19.6, are M.C.'s.

To return to (33), let us put, in analogy with the notation in Example 2, for any interval J in $\boldsymbol{T}$:

$$D_J = \prod_{r\in J}{}^{*} e_{q_r};$$

and for $r\in R$, $r'\in R$, $r\leqq r'$:

$$p_{rr'}(t) = D_{[r,r')}(t) - D_{[r,r']}(t). \tag{34}$$

For all other ordered pairs (s, s') in $\boldsymbol{T}$, we put $p_{ss'}(\cdot)\equiv 0$. The last term in (33) then becomes

$$\prod_{\nu=1}^{l} p_{r_{\nu-1}r_\nu}(t_\nu - t_{\nu-1}).$$

Thus $\{x_t, t\in \boldsymbol{T}\}$ is a M.C. with minimal state space R; essential range $\boldsymbol{T}$ (see the beginning of § 4); and transition matrix $(p_{rr'})$, $r, r'\in R$.

To verify that $q_{rr'}=0$ if $r<r'$, we observe that

$$D_{[r,r')}(t) = \int_0^t [1 - e^{-q_r(t-s)}]\, d D_{(r,r')}(s) \leqq q_r t D_{(r,r')}(t)$$

from which it follows that $\lim\limits_{t\to 0} D_{[r,r')}(t)/t = 0$ and consequently $\lim\limits_{t\to 0} (p_{rr'}t)/t = 0$ by (34). Therefore the Q-matrix of $(p_{rr'})$ has diagonal elements q_r but all other elements zero. This should be compared with Example 3 above; indeed if the only instantaneous state there could be eliminated the result should be an example of the type described here.

Example 8. The simplest example of a measurable but not standard transition matrix is given by (p_{ij}), $i, j\in \boldsymbol{I}$, where $\boldsymbol{I}$ contains N elements, $2\leqq N\leqq\infty$, and

$$p_{ij}(t)\equiv p_j>0, \quad i, j\in\boldsymbol{I}, \quad t\in\boldsymbol{T}^0; \quad \sum_{j\in\boldsymbol{I}} p_j = 1.$$

A M.C. $\{x_t, t\in\boldsymbol{T}\}$ with such a transition matrix is an *independent process*, namely in which all the x_t's are mutually independent random variables. Such a process, and any standard modification of it, is well-separable since each closed subset of $\bar{\boldsymbol{I}}$ is the intersection of a subfamily from a fixed denumerable, and for each $A\neq\boldsymbol{I}$ and open interval G with rational endpoints the set $\{\omega: x(t,\omega)\in A, t\in G\cap R\}$ in (4.10) is a null set so that the definition of separability is trivially satisfied. According to Theorem 4.3 then, there is no measurable standard modification since the transition matrix is not standard. This does not mean, however, that the sample functions $x(\cdot,\omega)$ cannot be all measurable or even Borel measurable. In fact, let $\tilde{\Omega}$ be the "function space" of all functions ξ with domain $\boldsymbol{T}$ and range $\bar{\boldsymbol{I}}$ and consider the probability measure pertaining to the independent process above, which assigns the measure $\prod_{\nu=1}^{n} p_{j_\nu}$ to the set $\{\xi(\cdot):\xi(t_\nu)=j_\nu,\ 1\leqq\nu\leqq n\}$. It is easy to see that the subset Ω of measurable functions (as well as its complement) in $\tilde{\Omega}$ has outer measure one. Hence by a well known procedure due to Doob [1; p. 68ff.] we may take Ω to be the probability space for the independent process and every sample function will be a measurable function.

If we form the direct product, as in Example 5, of the above (p_{ij}) with another standard transition matrix $(p_{i^*j^*})$ we obtain the situation described in (1.14) and (1.15). Each class J there is the set of all pairs (j^*, j) with a fixed j^*. The significance of Theorem 1.4 is now clear.

Example 9. Finally, we give an example of a transition matrix which is not measurable.

Lemma. Let $\boldsymbol{I}$ be a denumerable set of real numbers such that if i_1 and i_2 belong to $\boldsymbol{I}$ then $r_1 i_1+r_2 i_2$ belongs to $\boldsymbol{I}$ for every rational r_1 and r_2. There exists a function f from $\boldsymbol{T}^0$ to $\boldsymbol{I}$ such that

$$f(s+t)=f(s)+f(t).$$

Proof of Lemma. Let all the real numbers not in $\boldsymbol{I}$ be well-ordered as $\{\xi_\alpha\}$ and let ξ_1 be the first element. For an arbitrary set of real numbers S we define the *modul* $M(S)$ to be the set of all real numbers expressible as a finite linear combination of elements of S with rational coefficients. We construct a subset S_0 of $\{\xi_\alpha\}$ as follows. Let ξ_0 be an arbitrary element of $\boldsymbol{I}$. For a general transfinite ordinal $\alpha\geqq 1$, we put ξ_α in S_0 if and only if $\xi_\alpha\notin M(\{\xi_\beta,\ 0\leqq\beta<\alpha\})$. Thus in particular $\xi_1\in S_0$, and S_0 is defined by transfinite induction. Now let $\boldsymbol{M}=M(S_0)$. It is easy to verify that every real number can be expressed as a sum $i+v$ where $i\in\boldsymbol{I}$ and $v\in\boldsymbol{M}$, and the representation is unique. For every $t\in\boldsymbol{T}^0$ let $t=i_t+v_t$ be this representation and define $f(t)=i_t$. Clearly f maps $\boldsymbol{T}^0$ onto $\boldsymbol{I}$. Furthermore if $s=i_s+v_s$, then $s+t=(i_s+i_t)+(v_s+v_t)$; since $i_s+i_t\in\boldsymbol{I}$ and $v_s+v_t\in\boldsymbol{M}$ er have $f(s+t)=i_s+i_t=f(s)+f(t)$. □

Using the set $\boldsymbol{I}$ and the function f in the lemma, we define the elements of a matrix (p_{ij}), $i, j\in\boldsymbol{I}$, as follows:

$$p_{ij}(t)=\begin{cases} 1 & \text{if } f(t)=j-i;\\ 0 & \text{otherwise.}\end{cases}$$

To verify that this is a transition matrix we need only verify (1.B) and (1.C). To each t and i there corresponds a unique $j=i+f(t)$ for which $p_{ij}(t)=1$, hence (1.B) is true. Next $\sum_k p_{ik}(s)p_{kj}(t)=1$ or 0 according as there is or is not a k such that $f(s)=k-i$ and $f(t)=j-k$; namely according as $f(s+t)=f(s)+f(t)$ is or is not $j-i$. Hence (1.C) is true. For each i and j the function p_{ij} assumes both values 0 and 1 (since the mapping f is onto) and only these, hence it is not continuous and therefore by Theorem 1.1 not measurable.

If a M.C. has the above transition matrix then for every t we have $\boldsymbol{P}\{x(t,\omega)=i+f(t)\,|\,x(0,\omega)=i\}=1$.

Further examples of nonmeasurable transition matrices are obtained by forming the direct product of the above transition matrix with a measurable one.

Notes. We omit examples of *finite* M.C.'s and *birth and death processes*. There exists a considerable literature on each subject and more special methods are pertinent to its treatment.

Examples 2 and 3 are given by KOLMOGOROV [5] and analysed in detail by KENDALL and REUTER with semigroup methods [1]. The probabilistic analysis of Example 3 given here is new, confirming the inevitability of the construction at the end of Example 3 which is given by KENDALL and REUTER. Example 4 is a slight modification of one due to KENDALL [2]. The ramification given toward the end of Example 2 as well as examples similar to those discussed in Examples 1, 3 and 4 can be found in LÉVY [1]. Such examples are easily described and yet the rigorous verification that they yield M.C.'s with the required properties seems so laborious that one hopes for a more general and efficient method. Example 5 is an after-thought. Example 6 is due to BLACKWELL [4]. In LÉVY [1] it is incorrectly stated that not all states can be instantaneous. The first counterexamples were given independently by FELLER and MCKEAN [1] and DOBRUŠIN [3], but BLACKWELL'S example is the simplest. In the example of FELLER and MCKEAN the minimal state space is the set of rational numbers in [0, 1] and almost all sample functions are continuous, so that all the irrational numbers in [0, 1] are fictitious states and the essential range is the interval [0, 1]. For the analytic construction of a function which takes a given denumerable set of values a.e. without having any interval of constancy see Amer. Math. Monthly **66**, 428 (1959). Example 7 is due to LÉVY [1]. Simple as it may now seem, it behooves on him to perceive such a radical possibility. We have added some analytical details in this case which are typical in such constructions. REUTER [4] showed that there exists a (stochastic) transition matrix with a diagonal Q-matrix if and only if these diagonal elements are either unbounded or all zero. Example 8 is implicit in DOOB'S work on general concepts of stochastic processes. Example 9 is due to DOOB [1], slightly generalized by LÉVY [1].

Bibliography

AM = Ann. of Math.; AMS = Ann. Math. Statist.; PAMS = Proc. Amer. Math. Soc.; PNAS = Proc. Nat. Acad. Sci. U.S.A.; TAMS = Trans. Amer. Math. Soc.; TV = Teor. Veroyatnost. i Primenen; ZW = Z. Wahrscheinlichkeitsrechnung.

Austin, D. G.: [1] Some differentiation properties of Markoff transition probability functions. PAMS **7**, 756—761 (1956).

— [2] Note on differentiating Markoff transition functions with stable terminal states. Duke Math. J. **25**, 625—629 (1958).

— [3] A new proof of the strong Markov theorem of Chung. PNAS **44**, 575—578 (1958).

Blackwell, D.: [1] A renewal theorem. Duke Math. J. **15**, 145—150 (1948).

— [2] Extension of a renewal theorem. Pacific J. Math. **3**, 315—320 (1953).

— [3] On transient Markov processes with a countable number of states and stationary transition probabilities. AMS **26**, 654—658 (1955).

— [4] Another countable Markov process with only instantaneous states. AMS **29**, 313—316 (1958).

Blackwell, D., and D. Freedman: [1] The tail σ-field of a Markov chain and a theorem of Orey. AMS **35**, 1291—1295 (1964).

Blumenthal, R. M.: [1] An extended Markov property. TAMS **85**, 52—72 (1957).

Breiman, L.: [1] On transient Markov chains with application to uniqueness problem for Markov processes. AMS **28**, 499—503 (1957).

— [2] Transient atomic Markov chains with a denumerable number of states. AMS **29**, 212—218 (1958).

Chacon, R. V.: [1] Some theorems on continuous parameter Markov chains. Diss. Syracuse University 1956.

Chung, K. L.: [1] An ergodic theorem for stationary Markov chains with a countable number of states. Proc. Intern. Congr. Math. Cambridge, Mass. 1950, Vol. I, p. 568.

— [2] Contributions to the theory of Markov chains. J. Res. Nat. Bur. Stand. **50**, 203—208 (1953).

— [3] Contributions to the theory of Markov chains. II. TAMS **76**, 397—419 (1954).

— [4] Foundations of the theory of continuous parameter Markov chains. Proc. Third Berkeley Symposium on Math. Statist. and Probability, Vol. II, pp. 29—40. University of California Press 1956.

— [5] Some new developments in Markov chains. TAMS **81**, 195—210 (1956).

— [6] On a basic property of Markov chains. AM **68**, 126—149 (1958).

— [7] Some aspects of continuous parameter Markov chains. Publ. Inst. Statist., Univ. Paris **6**, 271—287 (1957).

— [8] On last exit times. Illinois J. Math. **4**, 629—639 (1960).

— [9] Probability methods in Markov chains. Proc. Fourth Berkeley Symposium on Math. Statist. and Probability, Vol. II, pp. 35—56. University of California Press 1961.

Chung, K. L.: [10] Some remarks on taboo probabilities. Illinois J. Math. **5**, 431—435 (1961).

— [11] On the Martin boundary for Markov chains. PNAS **48**, 963—968 (1962).

— [12] On the boundary theory for Markov chains. Acta Math. **110**, 19—77 (1963).

— [13] The general theory of Markov processes according to Doeblin. ZW **2**, 230—250 (1964).

— [14] On the boundary theory for Markov chains. II. Acta Math. **115**, 111—163 (1966).

Chung, K. L., and J. L. Doob: [1] Fields, optionality and measurability. Amer. J. Math. **87**, 397—424 (1964).

Chung, K. L., and P. Erdös: [1] Probability limit theorems assuming only the first moment. Mem. Amer. Math. Soc. No. 6, 1—19 (1951).

Chung, K. L., and W. H. J. Fuchs: [1] On the distribution of values of sums of random variables. Mem. Amer. Math. Soc. No. 6, 1—12 (1951).

Chung, K. L., and D. Ornstein: [1] On the recurrence of sums of random variables. Bull. Amer. Math. Soc. **68**, 30—32 (1962).

Chung, K. L., and J. Wolfowitz: [1] On a limit theorem in renewal theory. AM **55**, 1—6 (1952).

Derman, C.: [1] A solution to a set of fundamental equations in Markov chains. PAMS **5**, 332—334 (1954).

— [2] Some contributions to the theory of denumerable Markov chains. TAMS **79**, 541—555 (1955).

Dobrušin, R. L.: [1] Two limit theorems for the simplest random walk on a line. Uspehi Mat. Nauk. (N.S.) **10**, 139—146 (1955) [Russian].

— [2] On conditions of regularity of stationary Markov processes with a denumerable number of possible states. Uspehi Mat. Nauk (N.S.) **7**, 185—191 (1952) [Russian].

— [3] An example of a countable homogeneous Markov process all states of which are instantaneous. TV **1**, 481—485 (1956) [Russian].

— [4] Some classes of homogeneous denumerable Markov processes. TV **2**, 377—380 (1957) [Russian].

Doeblin, W.: [1] Sur les propriétés asymptotiques de mouvement régis par certains types de chaines simples. Bull. Math. Soc. Roum. Sci. **39**, No. 1, 57—115; No. 2, 3—61 (1937).

— [2] Sur l'équation matricielle $A^{(t+s)} = A^{(t)} A^{(s)}$ et ses applications aux probabilités en chaine. Bull. Sci. Math. (2), **62**, 21—32 (1938); **64**, 35—37 (1940).

— [3] Sur certains mouvements aléatories discontinus. Skand. Aktuarietidskr. **22**, 211—222 (1939).

— [4] Sur deux problèmes de M. Kolmogoroff concernant les chaines dénombrables. Bull. Soc. Math. France **66**, 210—220 (1938).

Doob, J. L.: [1] Topics in the theory of Markoff chains. TAMS **52**, 455—473 (1942).

— [2] Markoff chains — denumerable case. TAMS **58**, 455—473 (1945).

— [3] Renewal theory from the point of view of probability. TAMS **63**, 422—438 (1948).

— [4] Stochastic processes. New York 1953.

— [5] Discrete potential theory and boundaries. J. Math. Mech. **8**, 433—458 (1959).

Dynkin, E. B.: [1] Some limit theorems for sums of independent random variables with infinite mathematical exceptation. Izv. Akad. Nauk. SSSR., Ser. Mat. **19**, 247—266 (1955) [Russian].

DYNKIN, E. B.: [2] Foundations of the theory of Markov processes. Moscow 1959 [Russian].

DYNKIN, E. B., and A. A. YUŠKEVIČ: [1] Strong Markov processes. TV **1**, 149—155 (1956) [Russian].

ERDÖS, P., W. FELLER and H. POLLARD: [1] A theorem on power series. Bull. Amer. Math. Soc. **55**, 201—204 (1949).

ERDÖS, P., and M. KAC: [1] On certain limit theorems of the theory of probability. Bull. Amer. Math. Soc. **52**, 292—302 (1946).

FELLER, W.: [1] On the integro-differential equations of purely discontinuous Markoff processes. TAMS **48**, 488—575 (1940); Errata **58**, 474 (1954).

— [2] Fluctuation theory of recurrent events. TAMS **67**, 98—119 (1949).

— [3] An introduction to probability theory and its applications. New York, Vol. 1, 1950 (first ed.), 1957 (second ed.); Vol. 2, 1966.

— [4] Boundaries induced by stochastic matrices. TAMS **83**, 19—54 (1956).

— [5] On boundaries and lateral conditions for the Kolmogoroff differential equations. AM **65**, 527—570 (1957).

— [6] On the Fourier representation for Markov chains and the strong ratio theorem. J. Math. Mech. **15**, 274—283 (1966).

FELLER, W., and H. P. MCKEAN, Jr.: [1] A diffusion equivalent to a countable Markov chain. PNAS **42**, 351—354 (1956).

FOSTER, F. G.: [1] On Markoff chains with an enumerable infinity of states. Proc. Cambridge Phil. Soc. **47**, 587—591 (1952).

FRANK, P.: [1] Taboo generating functions and other topics in Markov chains. Diss. Columbia University 1959.

FRÉCHET, M.: [1] Recherches théoriques modernes sur le calcul des probabilitiés, Vol. II, Methode des fonctions arbitraires. Théorie des événements en chaine dans le cas d'un nombre fini d'états possibles. Paris 1938.

GNEDENKO, B. V., and A. N. KOLMOGOROFF: [1] Limit distributions for sums of independent random variables. Moscow-Leningrad 1949 (English translation by K. L. CHUNG. Cambridge, Mass. 1954).

HARDY, G. H.: [1] Divergent series. Oxford 1949.

HARRIS, T. E.: [1] First passage and recurrence distributions. TAMS **73**, 471—486 (1952).

— [2] Transient Markov chains with stationary measures. PAMS **8**, 937—942 (1957).

HARRIS, T. E., and H. ROBBINS: [1]: Ergodic theory of Markov chains admitting an infinite invariant measure. PNAS **39**, 860—864 (1953).

HARTMAN, P., and A. WINTNER: [1] On the law of the iterated logarithm. Amer. J. Math. **63**, 169—176 (1941).

HILLE, E., and R. S. PHILLIPS: [1] Functional analysis and semi-groups. Amer. Math. Soc. Colloq. Publ. 1957.

HODGES, J. L., and M. ROSENBLATT: [1] Recurrence-time moments in random walks. Pacific J. Math. **3**, 127—136 (1953).

HSU, P. L.: [1] The differentiability of the probability transition function of a purely discontinuous stationary Markoff process on the Euclidean space. Peking Univ. J. **4**, No. 2, 257—270 (1958).

HUNT, G. A.: [1] Markoff chains and Martin boundaries. Illinois J. Math. **4**, 313—340 (1960).

JAIN, N.: [1] Some limit theorems for a general Markov process. Diss. Stanford University 1965 (to appear in ZW).

JURKAT, W. B.: [1] On semi-groups of positive matrices, I. Scripta Math. **24**, 123—131 (1959).

— [2] On semi-groups of positive matrices, II. Scripta Math. **24**, 207—218 (1959).

— [3] On the analytic structure of semi-groups of positive matrices. Math. Z. **73**, 346—365 (1960).

KARLIN, S., and J. MCGREGOR: [1] Representation of a class of stochastic processes. PNAS **41**, 387—391 (1955).

KATO, T.: [1] On the semi-groups generated by KOLMOGOROFF's differential equations. J. Math. Soc. Japan **6**, 1—15 (1954).

KELLEY, J. L.: [1] General topology. New York 1955.

KEMENY, J., and L. J. SNELL: [1] Denumerable Markov chains. New York 1966.

KENDALL, D. G.: [1] Some analytical properties of continuous stationary Markov transition functions. TAMS **78**, 529—540 (1953).

— [2] Some further pathological examples in the theory of denumerable Markov processes. Quart. J. Math. **7**, 39—56 (1956).

— [3] A note on DOEBLIN's central limit theorem. PAMS **8**, 1037—1039 (1957).

— [4] Unitary dilations of one-parameter semigroups of Markov transition operators, and the corresponding integral representations for Markov process with a countable infinity of states. Proc. London Math. Soc. (3) **9**, 417—431 (1959).

— [5] Some recent developments in the theory of denumerable Markov processes (to appear).

KENDALL, D. G., and G. E. H. REUTER: [1] Some pathological Markov processes with a denumerable infinity of states and the associated semigroups of operators on *l*. Proc. Intern. Congr. Math. Amsterdam 1954, Vol. III, pp. 377—415.

— [2] The calculation of the ergodic projection for Markov chains and processes with a countable number of states. Acta Math. **97**, 103—144 (1957).

KINGMAN, J. F. C.: [1] An approach to the study of Markov processes. J. Roy. Statis. Soc. (13) (to appear).

KINGMAN, J. F. C., and S. OREY: [1] Ratio limit theorem for Markov chains. PAMS **15**, 907—910 (1964).

KOLMOGOROV, A. N.: [1] Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math. Ann. **104**, 415—458 (1931).

— [2] Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergebn. Math. **2**, No. 3, Berlin 1933.

— [3] Anfangsgründe der Theorie der Markoffschen Ketten mit unendlichen vielen möglichen Zuständen. Mat. Sbornik N.S. Ser. 607—610 (1936); Bull. Univ. Moscou **1** (1937) [Russian].

— [4] A local limit theorem for classical Markov chains. Izv. Akad. Nauk. SSSR., Ser. Mat. **13**, 281—300 (1949) [Russian].

— [5] On the differentiability of the transition probabilities in homogeneous Markov processes with a denumerable number of states. Učenye Zapiski MGY **148**, Mat. **4**, 53—59 (1951) [Russian].

KRENGEL, U.: [1] On the global limit behavior of Markov chains and of general nonsingular Markov processes (to appear in ZW).

LAMPERTI, J.: [1] An occupation time theorem for a class of stochastic processes. TAMS **88**, 380—387 (1958).

LÉVY, P.: [1] Systèmes markoviens et stationnaires. Cas dénombrable. Ann. Sci. Ecole Norm. Sup. (3) **68**, 327—381 (1951).

— [2] Complément à l'étude des processus de MARKOFF. Ann. Sci. Ecole Norm. Sup. (3) **69**, 203—212 (1952).

Lévy, P.: [3] Processus markoviens et stationnaires du cinquième type (infinité dénombrable des états possibles, paramètre continu). C. R. Acad. Sci. Paris **236**, 1630—1632 (1953).

— [4] Processus semi-markovien. Proc. Intern. Congr. Math. Amsterdam 1954, pp. 416—426.

— [5] Processus markoviens et stationnaires. Cas dénombrable. Ann. Inst. H. Poincaré **16**, 7—25 (1958).

Loève, M.: [1] Probability theory. 3rd ed. New York 1963.

Mauldon, J. G.: [1] On non-dissipative Markov chains. Proc. Cambridge Phil. Soc. **53**, 825—835 (1957).

Neveu, J.: [1] Une généralisation des processus à accroisements croissants indépendants. Abh. Math. Sem. Univ. Hamburg **25**, 36—61 (1961).

— [2] Lattice methods and submarkovian processes. Proc. Fourth Berkeley Symposium of Math. Statist. and Probability, Vol. II, pp. 347—391. University of California Press 1961.

— [3] Sur les états d'entrée et les états fictifs d'un processus de Markov. Ann. Inst. H. Poincaré, **17**, 324—337 (1962).

— [4] Chaines de Markov et théorie du potential. Ann. Fac. Sci. Claremont, No. 24, 37—89 (1964).

Orey, S.: [1] Sums arising in the theory of Markov chains. PAMS **12**, 847—856 (1961).

— [2] Strong ratio limit property. Bull. Amer. Math. Soc. **67**, 571—574 (1961).

— [3] An ergodic theorem for Markov chains. ZW **1**, 174—176 (1962).

Ornstein, D.: [1] The differentiability of transition functions. Bull. Amer. Math. Soc. **66**, 36—39 (1960).

Port, S. C.: [1] Ratio limit theorems for Markov chains. Pacific J. Math. **15**, 989—1017 (1965).

Ray, D.: [1] Resolvents, transition functions and strongly Markovian processes. AM **70**, 43—72 (1959).

Reuter, G. E. H.: [1] Über eine Volterrasche Integralgleichung mit totalmonotonem Kern. Arch. Math. **7**, 59—66 (1956).

— [2] Denumerable Markov processes and the associated contraction semi-groups on *l*. Acta Math. **97**, 1—46 (1957).

— [3] Denumerable Markov processes. II. J. London Math. Soc. **34**, 81—91 (1959).

— [4] Denumerable Markov processes. III. J. London Math. Soc. **37**, 63—73 (1962).

Reuter, G. E. H., and W. Ledermann: [1] On the differential equations for the transition probabilities of Markov processes with enumerably many states. Proc. Cambridge Phil. Soc. **49**, 247—262 (1953).

Romanovskii, V. I.: [1] Discrete Markov chains. Moscow-Leningrad 1949 [Russian].

Saks, S.: [1] Theory of the integral. Warzawa-Lwow 1937.

Sarymsakov, T. A.: [1] Elements of the theory of Markov processes. Moscow 1954 [Russian].

Smith, G. J.: [1] Instantaneous states of Markov processes. TAMS **110**, 185—195 (1964).

Smith, W. L.: [1] Asymptotic renewal theorems. Proc. Roy. Soc. Edinburgh (A) **64**, 9—48 (1954).

Spitzer, F.: [1] Principles of random walk. New York 1964.

Titchmarsh, E. C.: [1] The theory of functions. Oxford 1939 (second ed.).

Urbanik, K.: [1] Limit properties of homogeneous Markoff processes with a denumerable set of states. Bull. Acad. Polon. Sci. Cl. III **2**, 371—373 (1954).

Uspensky, U. V.: [1] Introduction to mathematical probability. New York 1937.

Veech, W.: [1] The necessity of Harris' condition for the existence of a stationary measure. PAMS **14**, 856—860 (1963).

Vere-Jones, D.: [1] Geometric ergodicity in denumerable Markov chains. Quart. J. Math. **13**, 17—28 (1962).

Widder, D. V.: [1] The Laplace transform. Princeton 1946.

Williams, D.: [1] A new method of approximation in Markov chain theory and its application to some problems in the theory of random time substitution. Proc. London Math. Soc. (3), **16**, 213—240 (1966).

— [2] The process extended to the boundary. ZW **2**, 332—339 (1963).

— [3] On the construction problem for Markov chains. ZW **3**, 227—246 (1964).

Yuškevič, A. A.: [1] On strong Markov processes. TV **2**, 187—213 (1957) [Russian].

— [2] On differentiability of transition probabilities of homogeneous Markov processes with a countable number of states. Yčenye Zapiski MGY **186**, Mat. **9** [Russian].

Index

Consult also the Contents for subject matter. References to authors and their works are found in the Notes at the end of sections and listed in the Bibliography. An asterisk indicates the word under which the item is listed. Some page numbers give keys rather than exact locations.

Abelian theorem 55
absolute distribution 9, 141
absorbing barrier 25
absorbing state 186
absorption 38
adjacent set 190
admissible family 167, 172
almost closed set 112
ascending escalator 274
atomic 116
augmented 1

Baire function 112
Baire's lemma 287
Bernoullian scheme 24
birth and death process 291
Borel-Cantelli lemma:
  If $\sum_n \boldsymbol{P}(\Lambda_n) < \infty$, then $\boldsymbol{P}(\limsup_n \Lambda_n) = 0$. If the Λ_n's are independent and $\sum_n \boldsymbol{P}(\Lambda_n) = \infty$, then $\boldsymbol{P}(\limsup_n \Lambda_n) = 1$.
Borel field ($=\sigma$-field) 1
—, generated 2
Borel measurable process 145
—, condition for 161, 165
Borel sets, linear 145
Boundary 118, 182, 212, 255, 263, 266, 271
—, crossing of 271
—, exits to and entrances from 266
—, Martin*

Cantor set, generalized 280
category 187, 219
central limit theorem 99, 109
—, counter-example to 107
—, local 111
chain with independent and stationary increments 10
—, special results on 51, 61, 92, 118
Chapman-Kolmogorov equation(s) 9, 119
class 13
class property 13
classification of states 13, 64, 183
  (see also under "positive state" and "recurrent")
closed set 16
communicate 12, 183
—, under taboo 48
compactified state space 145
complete measure 1
completely nonatomic 116
completion (of a measure) 145
conditional expectation 3, 165
conditional independence 178
conditional probability 2, 165
conditional probability distribution 168
conjugate system 39, 117
c.p. (= continuous parameter) 140
convolution 55
coordinate function 7
cyclical matrix 11
cylinder set 7

decomposition formulas 46
descending escalator 275
determining system 35, 52, 222
—, conjugate*
derivate (= Dini derivate) 135
Dini's theorem on derivates:
  Each of the four derivates and the difference quotients of a continuous function in an (open) interval have the same bounds.
Dini's theorem on uniform convergence:
  If a nondecreasing sequence of continuous functions converges to a continuous limit in a compact interval, then the convergence is uniform there.

discontinuity set 164
discrete skeleton 132, 183
discrete topology 144
dissection formula 83
distribution function 2
Doeblin's method 90, 111
domain of definition, of finiteness 2
Doob's construction 272
Doob's theorems (on separability and measurability) 144, 146
d. p. (= discrete parameter) 140

entrance law 266, 270
entrance time 81, 233
—, first*
ergodic theorem 34, 90, 92, 98
essential range 140, 144
essential state 13, 183
event (= set measurable $\mathscr{F}$) 1
expectation 3
exponential distribution 152, 229, 233, 276
—, memoryless property of 288
existence theorem (for Markov chains) 7, 141
exit time 233
—, first 187
—, last 199

Fatou's lemma:
If $f_n \geqq 0$ and ν is a measure, then $\varliminf_{n\to\infty} \int f_n\, d\nu \geqq \int \left(\varliminf_{n\to\infty} f_n\right) d\nu$.
Feller's construction (= minimal* solution)
fictitious state 140, 147, 255
fine topology 190
finite M.C. (= one with a finite state space) 9, 34, 111, 260, 291
finite-dimensional joint distributions 141
first entrance (= passage) time 187
first entrance time distribution 61, 193
—, analytical properties of 205, 237
—, dual to 201
—, mean of 51, 214
—, moments of 61
first entrance formula(s) 46, 195
first infinity 258
—, distribution of 255
first system (= backward equations) 246, 250
Fubini's theorem on differentiation 136
Fubini's theorem on product measures (see Saks [1; p. 82]): If $f(.,.)$ is integrable with respect to the product measure $\mu \times \nu$ or its completion, then $f(., v)$ is μ-integrable except for a set of v of ν-measure zero, $\int f(u, .)\, d\mu$ is ν-integrable and $\iint f(u, v)\, d(\mu \times \nu) = \int [\int f(u, v)\, d\mu]\, d\nu$.
function space 290
functional 83, 225

generating function 52

hitching 274
homogeneous (= "with stationary transition probabilities") 140
—, spatially 10
honest process 139

independent fields [processes]: Fields [processes] are independent iff any collection of sets [random variables], one from each field [process], are independent.
independent and stationary increments, chain* with
independent process 289
independent random variables, sets 3
indicator 3
inessential 13
infinite state sequence 112
initial distribution 6, 141
i.o. (= infinitely often)
instantaneous state 154
—, examples of 278, 285
—, special properties of 159, 163, 189, 190, 219
integral (of a process) 151, 225
inverse probabilities 54
invariant 112
isomorphism 113

jump 164
(see also under "pseudo-jump")
jump matrix 259

Kolmogorov's differential equations 245
Kolmogorov's extension theorem 7, 142
Kronecker's symbol 8

Laplace transform 213, 238
last exit decomposition 46, 201, 207

last exit time 199
law of the iterated logarithm 106
Lipschitz condition 135
lower semicontinuous 161

Markov chain (see also under "M.C.")
—, continuous parameter 140
—, discrete parameter 4
—, reverse*
Markov chain with stationary transition probabilities (= homogeneous Markov chain = M.C.) 6, 140
Markov property 4, 140, 197
—, strong*
M.C., construction of 267, 270, 274
—, finite*
—, indecomposable 17
—, irreducible 17
—, non-dissipative 39
—, open (on parameter set $T^0 = (0, \infty)$) 142, 275
—, reduced 15, 17
—, stopped 80, 272
—, versions of 161
—, with some class property 11
Martin boundary 267, 271
martingale 21, 41, 114, 118, 157, 191, 259, 265
mathematical expectation (= expectation)
matrix theory 9
mean recurrence time 28
measurable process 145
—, Borel*
—, progressively 162
measurable transition matrix 120
method of first entrance 19
metrically dense 152
minimal chain 254
minimal completion (of a stopped process) 273
minimal solution 253
—, uniqueness of 253, 270
minimal state space (= state space)

nonmeasurable process 289
nonmeasurable transition matrix 290
nonrecurrent 19, 186
nonstationary 8, 197
Nörlund mean 21
normal distribution 99
n-step transition matrix, probability 9

null set 1
null state 31, 187

one-step transition matrix, probability 6, 128
optional random variable 77, 167
—, reverse 198
optional starting [stopping] time 77, 79
optional sampling 230
order (of a state) 64

period 13, 183
point of density 152
point of jump*
Poisson process 260, 282
positive state 31, 187
possible value 2
post-α field 172
post-α process 78, 172
—, separability of 177
post-exit process 231
potential theory 118, 271
pre-α field 77, 167
probability measure, space, triple 1
product Borel field, measure, space 145
product M.C. 285
progressively measurable 162
pseudo-jump 243
—, from a stable state 236

Q-matrix, 134, 251
—, conservative 251
—, with bounded diagonal 260

random variable 1
—, denumerably-valued (= discrete) 140
—, finite-valued 2
random walk 12, 24, 28, 70, 71
ratio limit theorem 50, 54, 56, 70, 216, 221
—, individual 58, 61
recurrent 19, 181, 263
recurrent event 45
reduced probability triple 77
reflecting barrier 26
regular 40, 53
reverse chain 198, 265
return interval, time 81
renewal process 42, 240
renewal theory 45, 241
repetitive pattern 41

right continuity (of field family) 166
right lower* semicontinuous

sample function 140
—, discontinuity* set of
—, first [last] discontinuity of 248
sample point, space 1
semigroup 119, 128, 291
semi-Markovian 45
second system (= forward equations) 246, 250
separable 143
separability set 143
set of constancy 148
shift 112, 118, 239
sojourn set 115
sojourn time 233
standard transition matrix 128
standard modification 142
stable interval 154
stable state 154
—, special properties of 152, 153, 164, 233
state 4
—, fictitious*, instantaneous*, stable*
state space 4, 140
—, compactification of 145, 148, 182, 263, 266
stationary M.C. 36
stationary absolute distribution 38
—, generalized 53, 225
stationary transition probabilities 6
step function 258
—, condition for sample 260, 261
stochastically continuous 143
stochastic matrix 9, 119
stochastic process
—, continuous parameter 140
—, discrete parameter 3
—, measurable*, separable*, well-separable*
stopped process 272
stopping random variable 272 (see also under "optional stopping time")
strong law of large numbers 92, 96
strong Markov property 178, 181 (see also under "system theorem")
—, for entrance time 79, 181, 196, 227
—, for exit time 231
—, for time spent in a given state 229
strong ratio theorem (see "individual ratio* limit theorem")
subclass 14
substochastic transition matrix 54, 138
subsystem 246
substitution property 177
superregular 40, 53
system theorem 77

taboo probability, set 45, 188, 255
taboo probability function 188
— for post-exit process 237
Tauberian theorem 55, 56, 214
transfinite induction 272, 290
transient state (= nonrecurrent state)
transient set 113
transition matrix (one-step) 6, 128
transition matrix (function) 119, 128, 141
—, equicontinuity at zero of 129
—, measurable*, nonmeasurable*, standard*, substochastic*
—, omission of row condition of 133, 138, 212
transition probability function 141
—, asymptotic behavior of 33, 183, 216
—, behavior of derivative at 0 and ∞ 221, 212
—, continuity of 120, 123, 129
—, decomposition of 123
—, differential equations for 245
—, differentiation of 131, 132, 134, 139, 210, 235, 237, 239
—, inequalities for 246
—, integral representations of 201, 205
—, vanishing 126, 149, 234
—, via jumps 250
—, with taboo* set

uniformly integrable 95
unit matrix 9
unit distribution 240

variation 208
version of stochastic process 142
version of conditional expectation 165

weak law of large numbers 96, 110
well-separable 143

510.8 G889 V104 1967

a39001 006890126b

.8
89
104
1967

721

73-1